JOURNAL OF VIROLOGY, Apr. 2003, p. 3893–3897 Vol. 77, No. 7
0022-538X/03/$08.00+0 DOI: 10.1128/JVI.77.7.3893–3897.2003

MINIREVIEW

Molecular Clocks and the Puzzle of RNA Virus Origins

Edward C. Holmes*

Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

Although the ultimate origins of RNA viruses are uncertain, it seems reasonable to assume that these infectious agents have a long evolutionary history, appearing with, or perhaps before, the first cellular life-forms (38). While the RNA viruses we see today may not date back quite this far, the evidence that some DNA viruses have evolved with their vertebrate hosts over many millions of years (24) makes an equally ancient history for RNA viruses a natural expectation. Yet a very different picture of RNA virus origins is painted if their gene sequences are compared; by using the best estimates for rates of evolutionary change (nucleotide substitution) and assuming an approximate molecular clock (21, 33), it can be inferred that the families of RNA viruses circulating today could only have appeared very recently, probably not more than about 50,000 years ago. Hence, if evolutionary rates are accurate and relatively constant, present-day RNA viruses may have originated more recently than our own species.

Before discussing the solutions to this apparent paradox, it is important to determine exactly why the molecular clock estimates of RNA virus origins are so recent. The key to establishing a timescale of viral evolution lies in accurately determining the rate of nucleotide substitution. Most analyses undertaken to date suggest that the average rate of nucleotide substitution in RNA viruses is $\sim 10^{-3}$ substitutions per site per year, with an approximately fivefold range around this (21). The fact that broadly similar rates are found in RNA viruses with very different genome organizations and lifestyles implies that both the error rate associated with RNA polymerase, estimated to be about one mutation per genome replication (10), and the rate of viral replication are roughly constant. If the average substitution rate of $\sim 10^{-3}$ substitutions/site/year is accurate, then, on average, every nucleotide position will have fixed 1 substitution after ~1,000 years of evolution (corresponding to an average divergence time between two lineages of only 500 years). This also corresponds to an evolutionary (corrected) distance (d) between two sequences of 1.0, the maximum that can be reliably estimated through sequence comparisons; larger distances will be inherently inaccurate because of uncounted multiple substitutions at single sites. Of course, reality is a little more complex because viral genomes are a patchwork of synonymous sites, where mutations do not change the encoded amino acid, and nonsynonymous sites, where mutations alter amino acids and which usually evolve more slowly. If we conservatively assume that the substitution rate at nonsynonymous sites is roughly 100-fold less than that at synonymous sites, at $\sim 10^{-5}$ substitutions/site/year, then a distance d of 1 corresponds to a divergence time of ~50,000 years ago, although the greater influence of natural selection at nonsynonymous sites is likely to increase the variance around this. Hence, if two RNA viruses have an evolutionary distance of <1.0 at nonsynonymous sites, as is the case for many viruses classified in the same family and certainly for those within the same genus, then these viruses are unlikely to have diverged more than ~50,000 years ago if these substitution rates are accurate.

As a practical example, consider the genus *Flavivirus*, for which the evolutionary relationships among member viruses have been examined in detail. Phylogenetic trees of the flaviviruses contain three clades, corresponding to mosquito-borne viruses, tick-borne viruses, and viruses with no known vector (23). If synonymous (d_S) and nonsynonymous (d_N) distances are estimated among representative members of these clades for the NS5 gene (using the PAML package [46]; full results are available upon request), I find that the average d_S of the three groups is ~20 substitutions per site, indicating that synonymous sites are so saturated with substitutions that they cannot be reliably used to estimate evolutionary distances. However, the mean d_N of the three groups is roughly 100-fold less, at ~0.2 substitution per site, which would correspond to a divergence time of only ~10,000 years ago, assuming a nonsynonymous rate of 10^{-5} substitutions/site/year. Even if the nonsynonymous substitution rate is 10 times lower than this, the estimated divergence time would still only be 100,000 years ago, and to make the divergence of the flaviviruses correspond to the origin of the placental mammals at ~100 million years ago would require a nonsynonymous rate some 4 logs lower, at $\sim 10^{-9}$ substitutions/site/year!

Hence, simple molecular clock calculations suggest that the origin of RNA viruses is an extremely recent event. However, in some cases such a recent origin conflicts with other evolutionary data. Perhaps the most notorious example is that of the primate lentiviruses, which include the human immunodeficiency virus types 1 and 2 (HIV-1 and HIV-2) and a growing list of simian immunodeficiency viruses (SIVs) that infect a wide variety of African monkeys (19). At face value, it would appear that these viruses have been associated with their host species for millions of years. Not only are they asymptomatic in their natural hosts, which when compared to the high virulence of HIV suggests that they have evolved stable associations over an extended time period, but the phylogenies of the viruses and the hosts often match, which implies that the viruses and the hosts have undergone cospeciation. Although the divergence

* Mailing address: Department of Zoology, University of Oxford, South Parks Rd., Oxford OX1 3PS, United Kingdom. Phone: 44 1865 271282. Fax: 44 1865 310447. E-mail: Edward.Holmes@zoo.ox.ac.uk.

times of the primate species in question are often uncertain, it is clear that virus-host cospeciation must mean a viral evolutionary history dating back millions of years. In contrast, far more recent timings are revealed using the sort of molecular clock calculations described here. In this case, the maximum sequence divergence observed suggests that the deepest split among the primate lentiviruses occurred only a few thousand years ago, and clearly far more recently than their host species diverged (36).

A similar disparity between divergence times and phylogenetic history can be seen for another important human pathogen, hepatitis B virus (HBV). While not a strictly an RNA virus, HBV replicates using reverse transcriptase and so is also likely to evolve rapidly, even though precise estimates of substitution rates have proven difficult to obtain. Close relatives of HBV (the hepadnaviruses) naturally infect a variety of species of nonhuman primates, including chimpanzees, gorillas, orangutans, and gibbons (20, 43). Although the phylogeny of these viruses does not match those of their host species (12), it is intriguing that these primate species are those most closely related to humans, which hints that they have evolved together over a period of ~20 million years. Moreover, the primates that harbor hepadnaviruses live in very different geographical locations and it is difficult to see how these viruses could be so widely dispersed unless they had undergone cospeciation with their hosts (or were spread much more recently by humans). However, as with the primate lentiviruses, the best estimates of substitution rate suggest an evolutionary history of only a few thousand years and hence many orders of magnitude too recent for cospeciation (12).

Importantly, the discrepancy between phylogenetic evidence for long-term cospeciation and recent molecular clock dates is also seen for a number of single-stranded RNA viruses. There is some evidence that the phylogeny of isolates of GBV-C (hepatitis G virus) matches that of the human populations from which they are sampled (30). If true, this would mean that GBV-C has spread with its human hosts over a period of 100,000 years or so. More strikingly, the closest relative of GBV-C, the New World primate virus GBV-A, has seemingly undergone cospeciation with its host species over a period of a few million years (6). However, if this ancient evolutionary history for GBV-A/GBV-C is correct, then the overall substitution rate needs to be in the region of 10^{-7} to 10^{-8} substitutions/site/year (6, 39) and therefore far lower than those usually documented for RNA viruses. Moreover, high substitution rates in GBV-C (~10^{-4} substitutions/site/year) have also been observed in intrahost comparisons (27). A similar situation is apparent with the New World hantaviruses (family *Bunyaviridae*) and the arenaviruses, for which host and virus trees have been compared in detail. In most cases, it seems that these viruses have undergone cospeciation with their murid rodent hosts over several million years, although there is also evidence for occasional cross-species transfers (2, 3, 26, 31). Moreover, a history of cospeciation is not restricted to animal viruses and has also been proposed for the plant tobamoviruses (17). Whatever the virus, the problem is the same. If the inferred history of long-term virus-host cospeciation is correct, then the rates of nucleotide substitution in these viruses must be far lower than those reported for other RNA viruses; otherwise their gene sequences would be unrecognizably divergent.

WHY DO RNA VIRUSES APPEAR TO HAVE ORIGINATED SO RECENTLY?

At face value, there are three explanations for the apparent discrepancy between the very recent molecular clock estimates for RNA virus origins and the phylogenetic evidence for virus-host cospeciation over millions of years. The most obvious is that the molecular clock is not constant and that rates of nucleotide substitution have changed dramatically both between viruses and along lineages. Specifically, if some RNA viruses evolved much slower than others or have experienced periods when their rates of nucleotide substitution were reduced, then divergence times could be greatly extended.

An important example of such a change in substitution rate is provided by influenza A virus. Although synonymous substitution rates in influenza A viruses from aquatic birds, horses, pigs, and humans vary no more than is usual among RNA viruses, the nonsynonymous rate is greatly reduced for the avian viruses compared to that seen for human viruses (18). This supports a model in which there has been a substantial rate acceleration coinciding with the species jump from birds, in which influenza A virus persistently replicates in the gastrointestinal tract and is asymptomatic, to humans, in which it replicates in the respiratory tract, causes regular epidemics, and is subject to strong immune selection pressure (44). Similarly, the nonsynonymous substitution rate is lower in the asymptomatic SIVs than in the pathogenic HIVs (32). Whether such differences in nonsynonymous substitution rate are typical of RNA viruses that experience fundamental changes in their biology (persistence, tropism, virulence) as they switch hosts clearly requires further investigation.

While it is clear that nonsynonymous rates can vary to some extent when viruses infect new hosts, other than the debated case of GBV-C there is little evidence that overall substitution rates (that is, for synonymous and nonsynonymous sites combined) are orders of magnitude lower than 10^{-3} substitutions/site/year and hence sufficient to greatly extend viral divergence times. Indeed, the best evidence for low substitution rates is the supposed match between virus and host phylogenies, itself a subject of debate (see below). Such a lack of large-scale rate variation may not be unexpected. Because RNA polymerase has no repair activity, it is unlikely that intrinsic rates of mutation will vary substantially between RNA viruses or that they have been lower in the past than they are today. It is also unlikely that evolution could effectively grind to a halt if a virus reaches an adaptive peak, such as a stable association with a particular host species, so long as replication continues. Although the fixation rate of advantageous mutations might drop dramatically if there are few ways to improve fitness, which is seemingly the case for avian influenza viruses, many synonymous sites would still be expected to be neutral and therefore to accumulate sequence changes. For example, although the rate of nonsynonymous substitution is reduced in the asymptomatic SIVs, overall substitution rates are very similar to those seen for HIV (32). Furthermore, a continual accumulation of amino acid changes through time has been observed in a variety of RNA viruses (33), and experimental studies have provided little evidence for stasis in RNA virus evolution (28). In fact, this seems to be the rule in biological systems as a whole. As a case in point, living fossils like the coelacanth have

undergone very little morphological change over long periods of time, suggesting that they are at an adaptive peak, but have similar rates of nucleotide substitution to other vertebrates (48).

Perhaps the most likely way that clock rates could vary between RNA viruses is if there are major differences in rates of replication, that is, in virus generation times. For example, HIV and HBV both replicate using reverse transcriptase, but HBV undergoes many fewer replications per unit of time (29), which seems to have resulted in a lower substitution rate (12). Hence, if RNA viruses differ dramatically in replication rate or have experienced periods when replication was either latent or extremely slow, then this would reduce substitution rates in the long run and would correspondingly extend divergence times. There is at least one group of viruses, the human T-cell lymphotropic virus types 1 and 2 (HTLV-1/2), for which such a difference in replication rates has been demonstrated to have had a major effect on substitution rates. In epidemic situations where transmission rates are high, most notably in populations of injecting drug users, replication rates are rapid as the virus spreads quickly in each new host. This results in substitution rates similar to those seen for other RNA viruses (34). However, in regions of endemicity where vertical transmission is common, HTLV-1/2 maintain themselves within hosts through the clonal expansion of infected cells rather than through active replication. In this case, the substitution rate is reduced to that of the human DNA polymerases ($\sim 10^{-9}$ substitutions/site/year). The overall substitution rate for HTLV-1/2, estimated to be $\sim 10^{-6}$ substitutions/site/year (41), is therefore a composite of both long- and short-term rates. Finally, some RNA viruses establish persistent infections within their host species, which is also likely to result in lower rates of replication. As such, persistence might be expected to greatly reduce substitution rates in the long run and also to increase the likelihood of virus-host cospeciation (42), as perhaps is the case for the rodent hantaviruses.

A second explanation for the recent origin of RNA viruses is that the methods currently used to estimate evolutionary distances from gene sequences are flawed in some way, leading to a substantial underestimation of divergence times. Although there are a number of factors that influence the accuracy of these methods (for example, the probability of different types of base change and variation in base composition), the most likely source of error is in the relative rates of substitution for different sites along the sequence. If RNA viruses are biased such that substitution rates vary dramatically between sites, then this could have a major effect on distance estimates.

The development of methods that allow rates of nucleotide substitution to vary among sites has been one of the major advances in gene sequence analysis in recent years. The most commonly used method in this context utilizes a gamma distribution, in which a sequence alignment is divided into a number of classes, each with a different substitution rate. The precise distribution of these rate classes is then described by a shape parameter denoted α (45). Low α values (i.e., <1) mean that the sequence alignment is composed of both very quickly and very slowly evolving sites, and this appears to be true in most cases. Application of the gamma model can greatly inflate evolutionary distances and hence have a radical effect on divergence times. For example, an analysis of amino acid sequences from 57 proteins suggested that the divergence of eubacteria and eukaryotes only occurred ~ 2 billion years ago, some 1.5 billion years more recently than suggested by the fossil record (9). While this could be taken to mean that cellular life has a much more recent origin than is usually thought, a reanalysis of the same data, taking into account among-site rate variation using the gamma distribution, produced divergence times far closer to the 3.5 billion years that was expected (25). Although a 1.5-billion-year change in estimated divergence times is clearly a major effect, it is proportionally tiny compared to that required to reconcile molecular clock estimates of viral divergence times with those inferred under cospeciation.

Theoretically, however, it is possible to obtain ancient divergence times for RNA viruses simply by employing extremely skewed gamma distributions. Under this scenario, the vast majority of nonsynonymous substitutions take place at a very limited number of sites, while the rest of the sequence evolves far more slowly. As an example, consider the evolutionary distances estimated for second codon positions from our representative sample of flaviviruses (it is currently not possible to estimate gamma distances for nonsynonymous sites alone, so second positions serve as a rough approximation). For the three groups of flaviviruses, the mean d at these sites, corrected for multiple substitutions but without a gamma distribution, is ~ 0.25 and is similar to the nonsynonymous distance estimated previously. The maximum likelihood estimate for the shape parameter of the gamma distribution for these data is highly skewed ($\alpha = 0.34$). As expected, evolutionary distances increase if they are now estimated using this gamma model (mean $d = 0.43$), although not sufficiently to make a major difference to estimated divergence times, which only increase to a little over 20,000 years (again assuming a rate of 10^{-5} substitutions/site/year). However, more dramatic results are obtained if an even more skewed gamma distribution is used. If $\alpha = 0.1$, then d increases to 2.3, so that maximum divergence times will be in the region of 100,000 years ago. Likewise, $\alpha = 0.05$ equates to a mean distance of 52 substitutions per second codon site (maximum divergence time of ~ 2.5 million years ago), although the numbers are now so large that the distances cannot be considered accurate and the variation in distance among sites will be enormous. Similar observations have been made with respect to the primate lentiviruses (36).

An extreme bias in the rates of nucleotide substitution among sites can therefore have a dramatic effect on estimated divergence times. Could RNA viruses evolve in a way that among-site rate variation occurs in such a biased manner? There are reasons for thinking that such a scenario is realistic. RNA viruses clearly differ from other organisms in their remarkable capacity to mutate. An important evolutionary by-product of these high mutation rates is a cap on genome size; genomes larger than ~ 15 kb are rarely produced because of the "error threshold," the generation of a prohibitive number of deleterious mutations (11). Since viral genome sizes are limited, sequence regions will encode multiple functions and individual mutations will often have pleiotropic effects, such as those influencing both cell tropism and immune evasion (1). This, in turn, may mean that there are relatively few evolutionary pathways that can be followed by RNA viruses; otherwise, at least one key function will be disrupted, so that mutations preferentially accumulate at that small proportion of sites that are free to vary. Supportive evidence for such a model is the

frequency with which convergent evolution is observed for RNA viruses (4, 7, 13), as expected if only a limited number of evolutionary pathways are viable, and the evidence that RNA (37) and protein secondary structure (22) can act as constraints against sequence change. The problem, of course, is that the α values estimated using the analytical methods available at present are a long way from the extremely biased values required to produce such ancient divergence times. Whether these methods are systematically underestimating the extent of among-site rate bias in RNA viruses remains to be seen.

The final explanation for the recent origin of RNA viruses is that this is in fact what happened. The key to this hypothesis is that the molecular clock dates discussed throughout this paper only relate to the RNA viruses currently circulating, that is, those that have been identified over the last 100 years or so. Thus, RNA virus families like the flaviviruses may in fact have histories dating back many millions of years, but the early members of these families have gone extinct to be replaced by those we sample today. All that is left following these extinction events are the very long branches relating the different families of RNA viruses to each other. Although the absence of a fossil record for RNA viruses means that it is difficult to test this hypothesis directly, it is striking that comparisons of viruses from different families reveal extreme sequence divergence, such that they are often no more similar than random sequences would be (47). Indeed, the "tree" of all RNA viruses is highly distinctive in that it is composed of relatively close tips, representing members of each viral family, connected by internal branches of generally unquantifiable length. In these circumstances it is impossible to accurately infer when different families of RNA viruses diverged. As such, the lineages leading to the RNA virus families we see today could have existed for many millions of years but their early history has been erased by a combination of multiple substitution and continual extinction. Obviously, one implicit assumption in this model is that RNA viruses experience high rates of lineage birth (speciation) and death (extinction). Although unproven, the rapidity with which RNA viruses evolve within single populations makes it possible that they experience equally rapid macroevolution.

The recent origin of RNA viruses depicted by molecular clocks is therefore not as constraining as it seems. However, one major problem remains; how can we explain the match between virus and host phylogenies that has been taken as evidence for cospeciation over millions of years? In some cases, most notably the primate lentiviruses, a wider sampling of viruses has found so many exceptions to the strict match between host and virus trees that the validity of cospeciation has been questioned (5). Further, it is striking that for many other RNA viruses for which there is evidence for cospeciation, including GBV-A/GBV-C, arenaviruses, and hantaviruses, there are also clear examples of cross-species transmission. If these viruses evolve at the same rate as other RNA viruses, then a process other than cospeciation must have produced the remaining match between the virus and host trees. Two mechanisms seem most likely. First, if phylogenetically related host species tend to live sympatrically, then viruses will tend to jump between closely related host species, which appears to be the case for the hantaviruses (2). If this process occurs at sufficient frequency, then the host and virus trees may often match, giving a false impression of cospeciation. Second, the ability to jump species boundaries may be dependent on the phylogenetic distance between hosts, so that it is easier to establish a new infection in a closely related host species than in a more distantly related one (5, 8). This model at least has a veneer of biological reality as it is evident that as host gene sequences diverge, especially the cellular receptors to which viruses bind, the less likely it is that an invading virus will be able to infect a foreign cell type and establish a productive infection. Unfortunately, while both models are possible, they currently lack clear empirical support.

WHAT NEEDS TO BE DONE?

As this review has highlighted, there is currently no all-encompassing explanation for why the molecular clock dates of RNA virus origins are so recent. However, it is possible to identify areas where future research might be expected to clarify the situation. First, it is obviously important to obtain more accurate estimations of substitution rates in a wider array of RNA viruses. This can be most profitably achieved by examining viruses sampled over an extended time period, although such data can be difficult to collect. If RNA virus evolution is universally rapid, then we would expect to see a steady accumulation of mutations within the time frame of human observation. Conversely, viruses that pick up no or very few mutations during the sampling period would provide compelling evidence for slow clock rates. In the same way, it is clearly important to obtain better estimates of rates of viral replication in nature. In this context it is particularly important to determine whether persistent infections and viruses with low replication rates are also associated with low substitution rates, as this is perhaps the most likely way in which the molecular clock could differ between viruses. Indeed, it is possible that whether a virus causes an acute or a persistent infection represents a major division in the life-history strategy of viruses and that substitution rates might also vary accordingly (42).

Second, more studies of RNA virus evolution in animal and plant populations are required. If examples can be found of long-term virus-host cospeciation, without the exceptions that are visible with current data, then it will be difficult to not conclude that the viruses in question have an ancient evolutionary history. In turn, if the sequences of these viruses are not excessively divergent (i.e., d is <1 at nonsynonymous sites), then this would also constitute powerful evidence for a low rate of nucleotide substitution. Likewise, examining the extent to which nonsynonymous substitution rates change when viruses infect new host species is central to understanding the forces that affect the tick of RNA virus clocks.

Finally, it is clear that we need new models of sequence evolution that incorporate the idiosyncrasies of RNA virus evolution. There are a number of areas where improvements could be made. First, more sophisticated models of nonsynonymous evolution are required, specifically those that allow a limited number of sites to accumulate the vast majority of changes, as may be true of viruses in nature, as well as those that take account of RNA and protein secondary structure (15). At present, most attention has been directed toward gamma distribution, but it is possible that this model is too simplistic for the realities of RNA virus evolution and that more complex statistical models are required. It might also be

necessary to develop models that allow patterns and rates of nonsynonymous evolution to differ among lineages. Although this will clearly add to the complexity of models, it may represent an important step toward reality; in the case of the primate lentiviruses, amino acid residues that are highly variable in HIV are sometimes conserved in chimpanzee SIV (19), and this may have a significant effect on dating estimates. Such lineage-specific differences in rate might be incorporated using the so-called "covarion" model of sequence change, which has a long history in molecular evolution (15) and is currently experiencing a newfound popularity (14, 16). More radically, it might also be possible to develop epistatic models of sequence evolution, in which mutations that occur in one sequence region affect variability elsewhere in the genome. These models have some precedent in evolutionary genetics, for example in the guise of fluctuating neutral space (40), and may be a powerful innovation given the potential importance of compensatory changes in viral evolution. Although these developments are technically complex, they may greatly assist our attempts to accurately reconstruct the origin of RNA viruses.

REFERENCES

1. **Baranowski, E., C. M. Ruiz-Jarabo, and E. Domingo.** 2001. Evolution of cell recognition by viruses. Science **292:**1102–1105.
2. **Bohlman, M. C., S. P. Morzunov, J. Meissner, M. B. Taylor, K. Ishibashi, L. Rowe, S. Levis, D. Enria, and S. C. S. Jeor.** 2002. Analysis of hantavirus genetic diversity in Argentina: S segment-derived phylogeny. J. Virol. **76:** 3765–3773.
3. **Bowen, M. D., C. J. Peters, and S. T. Nichol.** 1997. Phylogenetic analysis of the *Arenaviridae*: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. Mol. Phylogenet. Evol. **8:**301–316.
4. **Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis, A. Gulati, C. Ho, and I. J. Molineux.** 1997. Exceptional convergent evolution in a virus. Genetics **147:**1497–1507.
5. **Charleston, M. A., and D. L. Robertson.** 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst. Biol. **51:**528–535.
6. **Charrel, R. N., P. De Micco, and X. de Lamballerie.** 1999. Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. J. Gen. Virol. **80:**2329–2335.
7. **Cuevas, J. M., S. F. Elena, and A. Moya.** 2002. Molecular basis of adaptive convergence in experimental populations of RNA viruses. Genetics **162:**533–542.
8. **DeFilippis, V. R., and L. P. Villarreal.** 2000. An introduction to the evolutionary ecology of viruses, p. 126–208. *In* C. J. Hurst (ed.), Viral ecology. Academic Press, San Diego, Calif.
9. **Doolittle, R. F., D. F. Feng, S. Tsang, G. Cho, and E. Little.** 1996. Determining the divergence times of the major kingdoms of living organisms with a protein clock. Science **271:**470–477.
10. **Drake, J. W., and J. J. Holland.** 1999. Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. USA **96:**13910–13913.
11. **Eigen, M.** 1987. New concepts for dealing with the evolution of nucleic acids. Cold Spring Harbor Symp. Quant. Biol. **52:**307–320.
12. **Fares, M. A., and E. C. Holmes.** 2002. A revised evolutionary history of hepatitis B virus (HBV). J. Mol. Evol. **54:**807–814.
13. **Fares, M. A., A. Moya, C. Escarmís, E. Baranowski, E. Domingo, and E. Barrio.** 2001. Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. Mol. Biol. Evol. **18:**10–21.
14. **Felsenstein, J.** 2001. Taking variation of evolutionary rates between sites into account in inferring phylogenies. J. Mol. Evol. **53:**447–455.
15. **Fitch, W. M.** 1971. Rate of change of concomitantly variable codons. J. Mol. Evol. **1:**84–96.
16. **Galtier, N.** 2001. Maximum-likelihood phylogenetic analysis under a covarion-like model. Mol. Biol. Evol. **18:**866–873.
17. **Gibbs, A. J., P. L. Keese, M. J. Gibbs, and F. García-Arenal.** 1999. Plant virus evolution: past, present and future, p. 263–285. *In* E. Domingo, R. Webster, and J. Holland (ed.), Origin and evolution of viruses. Academic Press, London, United Kingdom.
18. **Gorman, O. T., W. J. Bean, and R. G. Webster.** 1992. Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. Curr. Top. Microbiol. Immunol. **176:**75–97.
19. **Hahn, B. H., G. M. Shaw, K. M. de Cock, and P. M. Sharp.** 2000. AIDS as a zoonosis: scientific and public health implications. Science **287:**607–614.
20. **Hu, X., H. S. Margolis, R. H. Purcell, J. Ebert, and B. H. Robertson.** 2000. Identification of hepatitis B virus indigenous to chimpanzees. Proc. Natl. Acad. Sci. USA **97:**1661–1664.
21. **Jenkins, G. M., A. Rambaut, O. G. Pybus, and E. C. Holmes.** 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. **54:**152–161.
22. **Kelleher, A. D., C. Long, E. C. Holmes, R. L. Allen, J. Wilson, C. Conlon, C. Workman, S. Shaunak, K. Wulfestieg, P. Goulder, C. Brander, G. Ogg, J. S. Sullivan, W. Dyer, I. Jones, A. J. McMichael, S. Rowland-Jones, and R. E. Phillips.** 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27 restricted CTL responses. J. Exp. Med. **193:**375–385.
23. **Kuno, G., G.-J. J. Chang, K. R. Tsuchiya, N. Karabatsos, and C. B. Cropp.** 1998. Phylogeny of the genus *Flavivirus*. J. Virol. **72:**73–83.
24. **McGeoch, D. J., A. Dolan, and A. C. Ralph.** 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. J. Virol. **74:**10401–10406.
25. **Miyamoto, M. M., and W. M. Fitch.** 1996. Constraints on protein evolution and the age of the Eubacteria/Eukaryote split. Syst. Biol. **45:**568–575.
26. **Morzunov, S. P., J. E. Rowe, T. G. Ksiazek, C. J. Peters, S. C. St. Jeor, and S. T. Nichol.** 1998. Genetic analysis of the diversity and origin of hantaviruses in *Peromyscus leucopus* mice in North America. J. Virol. **72:**7–64.
27. **Nakao, H., H. Okamoto, M. Fukuda, F. Tsuda, T. Mitsui, K. Masuko, H. Lizuka, Y. Miyakawa, and M. Mayumi.** 1997. Mutation rate of GB virus C hepatitis G virus over the entire genome and in subgenomic regions. Virology **233:**43–50.
28. **Novella, I. S., C. L. Hershey, C. Escarmis, E. Domingo, and J. J. Holland.** 1999. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. J. Mol. Biol. **287:**459–465.
29. **Nowak, M. A., S. Bonhoeffer, A. M. Hill, R. Boehme, H. C. Thomas, and H. McDade.** 1996. Viral dynamics in hepatitis B virus infection. Proc. Natl. Acad. Sci. USA **93:**214–223.
30. **Pavesi, A.** 2001. Origin and evolution of GBV-C/hepatitis G virus and relationships with ancient human migrations. J. Mol. Evol. **53:**104–113.
31. **Plyusnin, A., and S. P. Morzunov.** 2001. Virus evolution and genetic diversity of hantaviruses and their rodent hosts. Curr. Top. Microbiol. Immunol. **256:**47–75.
32. **Rey-Cuillé, M.-A., J.-L. Berthier, M.-C. Bomsel-Demontoy, Y. Chaduc, L. Montagnier, A. G. Hovanessian, and L. A. Chakrabarti.** 1998. Simian immunodeficiency virus replicates to high levels in sooty mangabeys without inducing disease. J. Virol. **72:**3872–3886.
33. **Sala, M., and S. Wain-Hobson.** 2000. Are RNA viruses adapting or merely changing? J. Mol. Evol. **51:**12–20.
34. **Salemi, M., M. Lewis, J. F. Egan, W. W. Hall, J. Desmyter, and A. M. Vandamme.** 1999. Different population dynamics of human T cell lymphotropic virus type II in intravenous drug users compared with endemically infected tribes. Proc. Natl. Acad. Sci. USA **96:**13253–13258.
35. **Sanchez, A. J., K. D. Abbott, and S. T. Nichol.** 2001. Genetic identification and characterization of Limestone Canyon virus, a unique *Peromyscus*-borne hantavirus. Virology **286:**345–353.
36. **Sharp, P. M., E. Bailes, F. Gao, B. E. Beer, V. M. Hirsch, and B. H. Hahn.** 2000. Origins and evolution of AIDS viruses: estimating the time-scale. Biochem. Soc. Trans. **28:**275–282.
37. **Simmonds, P., and D. B. Smith.** 1999. Structural constraints on RNA virus evolution. J. Virol. **73:**5787–5794.
38. **Strauss, E. G., J. H. Strauss, and A. J. Levine.** 1996. Virus evolution, p. 141–159. *In* B. N. Fields, D. M. Knipe, and P. M. Howley (ed.), Fundamental virology, 3rd ed. Raven Press, New York, N.Y.
39. **Suzuki, Y., K. Katayama, S. Fukushi, T. Kageyama, A. Oya, H. Okamura, Y. Tanaka, Y. Mizokami, and T. Gojobori.** 1999. Slow evolutionary rate of GB virus C/hepatitis G virus. J. Mol. Evol. **48:**383–389.
40. **Takahata, N.** 1987. On the overdispersed clock. Genetics **116:**169–179.
41. **Van Dooren, S., M. Salemi, and A. M. Vandamme.** 2001. Dating the origin of the African human T-cell lymphotropic virus type-1 (HTLV-I) subtypes. Mol. Biol. Evol. **18:**661–672.
42. **Villarreal, L. P., V. R. DeFilippis, and K. A. Gottlieb.** 2000. Acute and persistent viral life strategies and their relationship to emerging diseases. Virology **272:**1–6.
43. **Warren, K. S., J. L. Heeney, R. A. Swan, Heriyanto, and E. J. Verschoor.** 1999. A new group of hepadnaviruses naturally infecting orangutans (*Pongo pygmaeus*). J. Virol. **73:**7860–7865.
44. **Webby, R. J., and R. G. Webster.** 2001. Emergence of influenza A viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. **356:**1817–1828.
45. **Yang, Z.** 1996. Among-site variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11:**367–371.
46. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biol. Sci. **13:**555–556.
47. **Zanotto, P. M. D. A., M. J. Gibbs, E. A. Gould, and E. C. Holmes.** 1996. A reassessment of the higher taxonomy of viruses based on RNA polymerases. J. Virol. **70:**6083–6096.
48. **Zardoya, R., and A. Meyer.** 1997. The complete DNA sequence of the mitochondrial genome of a "living fossil," the coelacanth (*Latimeria chalumnae*). Genetics **146:**995–1010.