HIV-1 Sequence Variation: Drift, Shift, and Attenuation

Minireview

Michael H. Malim*†§ and Michael Emerman‡§
*Department of Microbiology
†Department of Medicine
University of Pennsylvania School of Medicine
Philadelphia, Pennsylvania 19104
‡Division of Human Biology
Fred Hutchinson Cancer Research Center
Seattle, Washington 98109

The introduction and global dissemination of the retrovirus human immunodeficiency virus type-1 (HIV-1) in humans represents a dramatic and deadly example of recent genome emergence and expansion; since the beginning of the pandemic, over 50 million people have been infected and over 16 million of those have died of AIDS. As with all RNA viruses, HIV-1 replication is characterized by very high mutation rates. Technical advances made in DNA sequencing and the recovery of rare nucleic acids from diverse sources have facilitated the sequencing of HIV-1 on a massive scale. These studies have revealed the remarkable genome plasticity and continuing diversification of HIV-1 strains; the significance of this diversity for viral pathogenesis remains a major question in the field.

The replication of HIV-1, like that of all retroviruses, involves reverse transcription of the viral genome into a DNA copy that integrates into the host cell genome. As with other viral RNA replicases, reverse transcriptases lack proofreading capabilities owing to the absence of $3' \rightarrow 5'$ exonuclease activity. In the case of HIV-1, the misincorporation, deletion, insertion, or duplication of nucleotides occurs during reverse transcription with a frequency of 10^{-4} to 10^{-5}. This error frequency, coupled with an in vivo virus production rate exceeding 10^9 per day in an individual, the large number of infected people, and the persistent nature of infections, provides HIV-1 with tremendous scope for the generation of viral diversity (Figure 1a). For instance, circulating HIV-1 strains are continuing to diversify at about 0.0024 substitutions per base pair per year (Korber et al., 2000), a rate that is five to six orders of magnitude greater than for mammalian gene families.

Three distinct groups of HIV-1, called M (main), O (outlier), and N (non-M, non-O), have been identified and likely represent the products of three independent cross-species transmissions of a simian immunodeficiency virus that resides in chimpanzees (SIV$_{CPZ}$) into human populations (Hahn et al., 2000). Phylogenetic analyses further divide the M group, which make up the vast majority of circulating strains, into nine subtypes (A, B, C, D, F, G, H, J, and K) that are about 25% to 30% different from each other and share a single common ancestor that has been dated at around 1930 (1915–1941) (Korber et al., 2000). The "star-burst" phylogeny of the M group strains from this postulated root suggests

§E-mail: malim@mail.med.upenn.edu, memerman@fhcrc.org

that there is an exponentially growing population size (Peeters and Sharp, 2000) and highlights the evolutionary "success" of HIV-1 in humans.

A second mechanism for acquiring sequence diversity is recombination. This can occur when a cell that is dually infected with different viruses produces progeny virions with genomic RNAs from each virus, and strand-switching takes place during the next round of reverse transcription (Figure 1b). As increasing numbers of full-length viral sequences become available, the number of recombinant or mosaic viruses that are formed in this way from parental viruses of different subtypes is being recognized more frequently. Some of these recombinant genomes have themselves become established in the human population, and are classified as circulating recombinant forms (CRFs) (McCutchan, 2000; Peeters and Sharp, 2000). Importantly, and unlike the incremental accumulation of sequence changes that occurs through copying errors, recombination has the potential to introduce large numbers of genetic changes simultaneously.

We liken these two modes of achieving diversification to the processes of antigenic drift and antigenic shift that have been described for another RNA virus, influenza A virus. In this classical example of virus sequence variation, copying errors during genome replication result in minor sequence changes (antigenic drift), and reassortment of whole genome segments (the influenza genome comprises eight distinct segments) following coinfection with different viral serotypes yields viruses with novel segment combinations (antigenic shift). One fundamental difference between the influenza A virus and the HIV-1 subtype nomenclature is that influenza A subtypes correspond to virus phenotype (antigenicity) while assignment to an HIV-1 subtype is based on nucleotide sequence alone.

Different clades and CRFs are unequally distributed across the globe; for instance, subtype B dominates North America and Europe whereas A, C, and A/G recombinants are the most prevalent subtypes in Africa. A major question is whether HIV-1 subtypes segregate with any particular biological properties. Differences in host genetics (Rowland-Jones et al., 2001 [this issue of *Cell*]) and epidemiology within and between cohorts make the analyses of subtype-specific variations in viral phenotypes extremely difficult to sort out. Accordingly, evidence for biological differences between HIV-1 subtypes is scant, but nevertheless intriguing. For example, in one cohort where subtypes A, C, D, and G were all present, people infected with C strains had statistically lower CD4 T cell counts and higher viral loads (Neilson et al., 1999). Also, initial studies linked the presence of an additional NF-κB site in subtype C promoters to enhanced transcription and responsiveness to tumor necrosis factor-α in vitro, and, possibly, to increased replication in patients with chronic immune system activation. However, recent extensive sequencing efforts have shown that this extra site is only found in a minority of subtype C viruses (Rodenburg et al., 2001), thus highlighting the need to use multiple sequences to formulate hypotheses regarding any possible subtype-specific
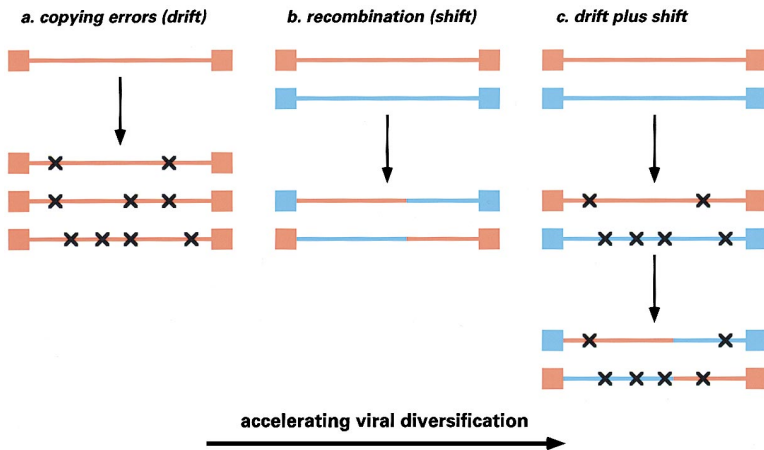
Figure 1. Modes of HIV-1 Sequence Diversification

The crosses signify copying errors made during reverse transcription, while red and blue proviruses represent phylogenetically distinct virus strains. Refer to the text for further details.

biology. HIV-1 and HIV-2 differ from one another in terms of pathogenicity, and by ~50% in nucleotide sequence. Although an imperfect comparison for HIV-1 subtype differences, such differences do point out the potential for biological variation between human lentivirus infections and disease progression.

Genetic Shift

There is ample evidence that genetic shift between retroviruses has occurred in ways that have helped shape the evolutionary history of viral populations. For instance, the SIV that infects the sabaeus species of African green monkeys in west Africa (SIV$_{AGM/SAB}$) is actually a complex recombinant between diverse SIV lineages found in different genera of African monkeys (Hahn et al., 2000). This phenomenon is not unique to members of the lentiviral genus of retroviruses, and is not even restricted to viruses that normally infect similar animal species. For instance, the reticuloendotheliosis virus family of oncoretroviruses, which infects turkeys, ducks, and other gallinaceous birds, has a recombinant genome that is similar to murine retroviruses at its 5′-end and to a class of primate retroviruses at its 3′-end (Kewal-Iramani et al., 1992). Thus, recombination has the potential to expand the host range of circulating viruses.

Since recombination between different retroviruses requires a common host, one can use the existence of recombinants as evidence that dual infections must have occurred. For example, the N group of HIV-1 (which has only been found in Cameroon) is a mosaic formed between divergent viral lineages most closely related to HIV-1 and SIV$_{CPZ}$. For coinfection of chimpanzees by the N group parental strains to have occurred, these two ancestral viruses must have had overlapping patterns of geographic distribution and a degree of shared tropism. Such lines of reasoning further support the argument that HIV-1 comprises viral lineages that are part of a larger family of viruses found in chimpanzees, and is, therefore, the product of SIV$_{CPZ}$ zoonoses (Hahn et al., 2000). Notably, there is still no direct evidence that these ancient recombination events conferred advantageous phenotypes other than the prevalence of the strains themselves.

With the HIV-1 M group viruses, recombination provides a major mechanism for generating "new" viruses and is occurring in "real-time." Specifically, the frequency of recombinants in cohorts where multiple clades are present has been estimated to be around 20% (Neilson et al., 1999). This level of abundance, together with indications that CRFs can replace existing subtypes in some populations (for example, a subtype A/G recombinant has become the predominant strain in parts of western and central Africa), suggests that certain recombinants may be able to spread faster than their parent subtypes (McCutchan, 2000). Although the supplanting of one viral subtype with another within a study population must be interpreted with caution because of the possible influence of "founder" effects, studying phenotypic variations between recombinants and their matching parental strains offers opportunities to address determinants of altered pathogenicity or transmission. For example, an HIV-1 A/B recombinant that is responsible for a rapid epidemic among intravenous drug users in Kaliningrad (Russia) is a recombinant in the *nef* gene and could shed some light on the in vivo significance of the multiple properties that have been attributed to this gene.

It was once thought that recombination is a relatively rare event since coinfection is limited by a phenomenon called "super-infection immunity" whereby one retrovirus generally precludes subsequent infection by a second retrovirus due to downregulation of the viral receptor. However, recent data suggests that coinfection of cells with HIV-1 may be common in vivo (Gratton et al., 2000). Thus, within a single host, recombination could be a major force for the rapid dissemination of multidrug resistance (see below), or for the redistribution of viral genes under selection pressure on the backbone of an already drug-resistant genome (Figure 1c).

Diversity during Disease Progression

High viral loads and host cell turnover characterize HIV-1 infection. Accordingly, the virus replicates as a heterogeneous and dynamic mixture of sequence variants, called a quasispecies, that continues to evolve for the duration of an infection. While the rate and extent of diversification differs from one infected person to the next, a recent longitudinal analysis of *env* gene variation determined that divergence from the founder (transmitted) strain, as well as overall diversity, both increased by ~1% per year in the early phases of typical infections (Shankarappa et al., 1999). At any given time, viral popu-

lations will be dominated by those strains that are most fit at that time. Because the selective pressures that determine which viral attributes confer fitness are constantly changing, the continual generation of diversity is presumably important for viral persistence in the infected host. This being said, the relative contributions of selection versus stochastic processes in driving virus diversification remain controversial and unresolved (Crandall et al., 1999).

One area of current uncertainty is whether disease progression is accompanied by the emergence of strains that are more pathogenic. Using the SIV/macaque experimental model for primate immunodeficiency virus infections, a number of studies have shown that serial passage of SIV (and HIV/SIV chimeras) from one animal to the next generates viruses that replicate with increased titers and/or display heightened virulence. In a related study, monkeys were challenged with cloned SIVs isolated during the early, mid, and late stages of an experimental infection; again, the late-stage virus grew to higher titers and induced more rapid progression to an AIDS-like disease (Kimata et al., 1999).

Many analyses have demonstrated that HIV-1 tropism evolves during infection: transmitted (early) strains utilize the CCR5 chemokine receptor for viral entry (R5 viruses), whereas many late-stage strains utilize CXCR4 instead (X4 viruses). However, while the appearance of X4 viruses frequently correlates with faster disease progression, it remains unclear whether this phenotypic switch is a cause or a consequence of immune system dysfunction. Moreover, the switch is not observed in all patients who progress to AIDS and is rare in people infected with subtype C viruses (McCutchan, 2000). It has frequently been suggested that X4 viruses are more fit than R5 viruses; in particular, single strain challenges in culture often show that X4 viruses replicate more rapidly and with greater cytopathicity. Recent studies, however, have begun to address the question of fitness by systematically evaluating replicative capacity by challenging cultures with mixtures of viruses and determining the "winner" strain over time (Quinones-Mateu et al., 2000). A number of interesting findings are emerging: (1) there is no evident correlation between replication fitness in vitro and the R5/X4 phenotype; (2) there is a correlation between fitness in culture and viral load in vivo; and (3) patients who are long-term survivors tend to be infected with viruses that are intrinsically less fit in culture. It will now be instructive to use this competition approach to assess whether (or how) replicative capacity changes during the course of HIV-1 infection.

In contrast to the observations made with SIV models (cited above), there is no evidence that HIV-1 increases in virulence as it is transmitted from one person to the next. What, then, are the possible explanations for the dichotomy between experimental SIV infections and natural HIV-1 infections in terms of the relationship between virus diversity and pathogenicity? The answer most likely relates to the process of transmission where selection pressures differ from those encountered within a single infected host. In serial passage experiments, such as those discussed for SIV, the use of bulk inocula containing multiple strains usually results in the rapid selection of viruses with augmented virulence, while nat-

ural transmission may select against overt virulence (Ebert, 1998). For example, it is well documented that new transmissions reestablish the selection for R5 viruses even if the index case contained both X4 and R5 viruses. Our current understanding of virulence and transmission is very incomplete, so it is likely that other viral determinants are also selected for or against during natural transmission. In addition, HIV-1 transmission is an inefficient process; for instance, less than 1% of heterosexual exposures result in successful transmission. This argues that few viral particles are capable of establishing a new infection. It is possible, therefore, that such transmission selects for particular strains that would constitute only a minority fraction of the challenge inoculum and may be relatively less fit in the context of an established infection.

A key selective pressure that helps propel viral diversity during the course of a single infection is evasion of the host immune response. Studies using the SIV system indicate that the virus continually adapts to evade the inhibitory action of neutralizing antibodies. That the humoral response helps to control SIV replication is further supported by the noted inverse correlations between viral virulence and sensitivity to neutralizing antibodies. For HIV-1, the evidence for the development of effective neutralizing antibodies during infection is much less convincing. Indeed, it has been argued that viral escape from such ineffective responses may be relatively insignificant compared to other selective processes, or that it may even be a byproduct of viral variation (Parren et al., 1999).

Though it has become clear that cytotoxic T lymphocytes (CTLs) play the predominant role in virus clearance during the acute phase of HIV-1/SIV infections, a definitive demonstration of CTL escape through sequence variation has been obtained only recently. Using the SIV model, Allen and colleagues (2000) measured CTL responses and sequenced circulating viruses for eight weeks following challenge. Importantly, the diminution of a strong CTL response against a Tat-specific epitope was accompanied by frequent changes in the corresponding region of the tat gene (82% of clones were altered by eight weeks post-infection). Further analysis of these samples also revealed that increased rates of sequence variation in this epitope were correlative with lower viral loads, a conclusion that is consistent with an earlier longitudinal study of HIV-1 infected subjects where increased rates of diversification tended to correlate with slower disease progression (Wolinsky et al., 1996). In sum, these findings suggest that while CTL escape and increased viral diversity may be necessary for averting virus elimination, such changes may cost the virus in terms of replicative potential by compromising important viral functions.

One unambiguous illustration of HIV-1's capacity to adapt in response to selection is the emergence of drug-resistant strains in patients receiving antiretroviral therapy. For example, treatment of HIV-1 infected people with single inhibitors of viral enzymes typically selects for resistant strains that predominate within weeks of the initiation of therapy. Even though combination therapies using multiple reverse transcriptase and protease inhibitors commonly suppress virus replication for prolonged periods, multi-drug resistant strains can evolve during

the course of infection. In addition to copying errors made during DNA synthesis, it is likely that recombination would also help to accelerate the development of such problematic strains (Figure 1c). Importantly, the drug-resistant strains that initially arise are typically less fit; over time, however, additional mutations can accumulate that increase fitness again. In a similar vein, it will now be interesting to determine if CTL escape variants are also intrinsically less fit. Whether the reduced viral loads that parallel greater diversity are attributable to more effective host responses and/or reductions in viral fitness remains to be determined.

Attenuation

Toward the low-end of the sliding scale of virus fitness is the generation of viruses with attenuated phenotypes. Such viruses are expected to exist in any virus population due to the high mutation rate of the virus. In an infected person with high viral burden, such viruses would normally be out-competed by the descendents of viruses with greater fitness. However, attenuated viruses could occasionally become fixed if the virus passes through an extreme "bottleneck" as low numbers of infectious particles establish new transmissions (Duarte et al., 1992). Studies of HIV-1 infected people who progress to disease very slowly or not at all (called long-term nonprogressors or survivors) show that a minority of these cases are, in fact, attributable to obviously attenuated viruses. One of the most prominent examples of this is a cohort of transfusion recipients in Sydney who were infected with a virus that lacked an intact nef gene before 1985, and who have remained mostly asymptomatic (Learmont et al., 1999). It is likely that continued analyses of full-length virus sequences will identify additional cases; studying the defects in these viruses has the potential to be very revealing of viral functions that are important in vivo.

It is important to reemphasize that viruses do not necessarily evolve toward greater pathogenicity. Indeed, the SIVs that naturally infect non-human primates in Africa discussed above do not cause observable disease in their natural hosts, yet can clearly induce fatal diseases when transferred to other species. Since some of these SIVs have resided in their host species for (at least) thousands of years, they might be examples of viruses that have become attenuated for disease in their natural hosts. There are numerous examples of viruses that are ubiquitous in the human population that are usually not pathogenic. Given time, is it possible that HIV-1 might similarly evolve? In essence, we do not know if HIV-1 diversification is yielding more, less, or equally pathogenic strains. Only difficult studies that combine extensive sequence data, standardized natural history, and carefully measured biological determinants will be able to begin to answer this question.

Selected Reading

Allen, T.M., O'Connor, D.H., Jing, P., Dzuris, J.L., Mothe, B.R., Vogel, T.U., Dunphy, E., Liebl, M.E., Emerson, C., Wilson, N., et al. (2000). Nature 407, 386–390.

Crandall, K.A., Vasco, D.A., Posada, D., and Imamichi, H. (1999). AIDS 13, S39–S47.

Duarte, E., Clarke, D., Moya, A., Domingo, E., and Holland, J. (1992). Proc. Natl. Acad. Sci. USA 89, 6015–6019.

Ebert, D. (1998). Science 282, 1432–1435.

Gratton, S., Cheynier, R., Dumaurier, M.-J., Oksenhendler, E., and Wain-Hobson, S. (2000). Proc. Natl. Acad. Sci. USA 97, 14566–14571.

Hahn, B.H., Shaw, G.M., De Cock, K.M., and Sharp, P.M. (2000). Science 287, 607–614.

Kewalramani, V.N., Panganiban, A.T., and Emerman, M. (1992). J. Virol. 66, 3026–3031.

Kimata, J.T., Kuller, L., Anderson, D.B., Dailey, P., and Overbaugh, J. (1999). Nat. Med. 5, 535–541.

Korber, B., Muldoon, M., Theiler, J., Gao, F., Gupta, R., Lapedes, A., Hahn, B.H., Wolinsky, S., and Bhattacharya, T. (2000). Science 288, 1789–1796.

Learmont, J.C., Geczy, A.F., Mills, J., Ashton, L.J., Raynes-Greenow, C.H., Garsia, R.J., Dyer, W.B., McIntyre, L., Oelrichs, R.B., Rhodes, D.I., et al. (1999). New Engl. J. Med. 340, 1715–1722.

McCutchan, F.E. (2000). AIDS 14, S31–S44.

Neilson, J.R., John, G.C., Carr, J.K., Lewis, P., Kreiss, J.K., Jackson, S., Nduati, R.W., Mbori-Ngacha, D., Panteleeff, D.D., Bodrug, S., et al. (1999). J. Virol. 73, 4393–4403.

Parren, P.W., Moore, J.P., Burton, D.R., and Sattentau, Q.J. (1999). AIDS 13, S137–S162.

Peeters, M., and Sharp, P.M. (2000). AIDS 14, S129–S140.

Quinones-Mateu, M.E., Ball, S.C., Marozsan, A.J., Torre, V.S., Albright, J.L., Vanham, G., van Der Groen, G., Colebunders, R.L., and Arts, E.J. (2000). J. Virol. 74, 9222–9233.

Rodenburg, C.M., Li, Y., Trask, S.A., Chen, Y.J., Robertson, D.L., Kalish, M.L., Shaw, G.M., Allen, S., Hahn, B.H., Gao, F., et al. (2001). AIDS Res. Hum. Retroviruses 17, 161–168.

Rowland-Jones, S., Pinheiro, S., and Kaul, R. (2001). Cell 104, this issue, 473–476.

Shankarappa, R., Margolick, J.B., Gange, S.J., Rodrigo, A.G., Upchurch, D., Farzadegan, H., Gupta, P., Rinaldo, C.R., Learn, G.H., He, X., et al. (1999). J. Virol. 73, 10489–10502.

Wolinsky, S.M., Korber, B.T., Neumann, A.U., Daniels, M., Kunstman, K.J., Whetsell, A.J., Furtado, M.R., Cao, Y., Ho, D.D., and Safrit, J.T. (1996). Science 272, 537–542.