

# Selection Forces and Constraints on Retroviral Sequence Variation

Julie Overbaugh<sup>1</sup> and Charles R. M. Bangham<sup>2</sup>

All retroviruses possess a highly error-prone reverse transcriptase, but the extent of the consequent sequence diversity and the rate of evolution differ greatly among retroviruses. Because of the high mutability of retroviruses, it is not the generation of new viral variants that limits the extent of diversity and the rate of evolution of retroviruses, but rather the selection forces that act on these variants. Here, we suggest that two selection forces—the immune response and the limited availability of appropriate target cells during transmission and persistence—are chiefly responsible for the observed sequence diversity in untreated retroviral infections. We illustrate these aspects of positive selection by reference to specific lentiviruses [human and simian immunodeficiency viruses (HIV and SIV)] and oncoviruses [feline leukemia virus (FeLV) and human T cell leukemia virus (HTLV)] that differ in their extent of variation and in disease outcomes.

Retroviruses acquire a point mutation on average once every replication cycle because the viral polymerase, reverse transcriptase, cannot correct nucleotide misincorporation errors. The genetic complexity of the viral population in the host will therefore be determined, in part, by the number of rounds of replication that occur. Retroviruses, such as HIV, that replicate continuously and at a high level (1, 2) therefore develop an extraordinary degree of sequence diversity in each infected host: A typical untreated HIV-1-infected person is likely to possess HIV-1 genomes with every possible single-base error. Indeed, the error rate of retroviral reverse transcriptases ( $10^{-4}$  to  $10^{-5}$  substitutions per site per replication cycle) appears to be close to the maximum possible for the genome size (<10 kb) typical of a small virus; if the error rate were higher, then deleterious mutations would be too frequent to allow the virus to survive. Even accounting for the error generation rate of reverse transcriptase, negative selection against less fit viruses is likely to be a dominant force.

Positive selection also influences the sequence diversity of retroviruses. This is perhaps most clearly illustrated by immune selection, which leads to viral escape from specific host defenses. In addition, for a retrovirus to maintain a persistent infection, there may be positive selection for viruses that can adapt to find new target cells in the host. Positive and negative selection forces can differ both between individual hosts and at different stages of the virus's life history: Rapid sequence diversification in one host does not necessarily lead to rapid evolution of the virus in the population if there are selective constraints that act on viral transmis-

sion to a new host. The intensity of selection can be quantified by the extent to which nonsynonymous nucleotide mutations (those that change an amino acid residue) occur more often than would be expected by chance [the Dn/Ds ratio (3)].

The error rates of different reverse transcriptase enzymes appear to differ by a factor of 4 or less (4), yet there are great differences in sequence diversity between different retroviruses. Below, we illustrate the different patterns of selection that act on retroviruses by considering specific examples: two oncogenic retroviruses, FeLV and HTLV-1, and the lentiviruses HIV-1 and SIV.

## Feline Leukemia Virus

Feline leukemia virus (FeLV) is considered an oncogenic retrovirus because it causes cell proliferation leading to leukemia and lymphoma. But there are variant FeLV strains that cause cytopathic effects that result in immunodeficiency disease (5). Studies in the FeLV system have shown that specific viral genotypes and phenotypes are closely linked with the variable disease outcomes associated with infection (5). This model system also clearly illustrates the principle that the viruses that are selected for their fitness for transmission from host to host may differ from those that are selected during persistent infection. There are four subgroups of FeLV (A, B, C, and T) that are distinguished by the receptors they require for entry into the cell (6). One subgroup, FeLV-A, is transmitted from host to host (7), presumably because cells that express the receptor for this virus are important target cells at the site of initial infection or in early virus amplification in the host.

FeLV-A is highly conserved, with less than 2% amino acid difference in the envelope glycoprotein surface unit, which is the receptor binding protein, between FeLV-A isolated more than a decade apart from cats living on

different continents (8). Other FeLV subgroups evolve through point mutations, small insertions, and recombination with cellular sequences (5). Specific scattered amino acid changes and a small insertion in the envelope protein have resulted in the emergence of a cytopathic variant, FeLV-T, which causes lymphoid depletion and immunodeficiency disease (5, 9). The cell-type specificity of FeLV-T arises because a cofactor that is required for entry by FeLV-T, but not by other FeLVs, is most highly expressed in lymphoid cells (6).

In all exogenous retroviruses, there is frequent recombination between the two copies of the viral genome present in the particle. Such recombination is thought to eliminate harmful mutations and to allow the virus to reconstitute a viable genome (10). In the case of FeLV infection, recombination also occurs during reverse transcription between FeLV-A genomes and related endogenous retroviral elements, resulting in new FeLV strains known as FeLV-B (11). For FeLV and other simple animal retroviruses, recombination with cellular sequences leads to a change in receptor specificity. Interestingly, the cellular receptor for both the FeLV-B and FeLV-T variants and for some murine and primate oncogenic retroviruses is a phosphate transporter molecule (6, 12–14). The evolution of several retroviruses to use similar receptors suggests that these proteins, and the cells that express them, offer advantages for virus replication during chronic infection. Indeed, the phosphate transporters are widely expressed in many cells and tissues (15); moreover, they are proteins that not only permit binding of retroviral envelope proteins, but also are competent to carry out other stages in viral entry, such as fusion (16). This requirement for a surface molecule to participate in both binding and fusion might limit the repertoire of cellular proteins that the virus can use as receptors.

## Human T Cell Leukemia Virus

The human T cell leukemia virus-type 1 (HTLV-1) is a more complex oncoretrovirus than FeLV; in addition to the *gag*, *pol*, and *env* genes also found in FeLV, HTLV includes genes that encode viral regulatory proteins. However, HTLV-1 resembles FeLV in the relatively low sequence diversity seen in natural isolates (17–19). Only 1 to 4% nucleotide sequence difference occurs among isolates from different continents, although there are distinct geographical distributions of minor sequence variants (19). Disease is less common in HTLV-1 infection than in FeLV or most other

<sup>1</sup>Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. E-mail: joverbau@fhcrc.org <sup>2</sup>Imperial College School of Medicine, London W2 1PG, UK. E-mail: c.bangham@ic.ac.uk

retroviral infections; it occurs in only 5% of infected individuals. About 2 to 3% develop a chronic inflammatory disease such as HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP), and 1 to 2% develop an aggressive T cell malignancy (20).

HTLV-1 is also relatively invariant in sequence within the host, although minor sequence variants are common (18). There is recent evidence that the nucleotide misincorporation rate of HTLV-1 reverse transcriptase is significantly lower than that of other viral reverse transcriptases (5), but the fidelity is not sufficiently high to explain the observed low degree of sequence diversity of HTLV-1. This suggests that HTLV-1 undergoes few complete rounds of replication in each host, and that the very high proviral loads frequently found in HTLV-1 infection (21) are maintained chiefly by mitosis of HTLV-1-infected T lymphocytes (22). Thus, the HTLV-1 genome appears to be replicated mainly by a cellular DNA polymerase, which has a much lower error rate than that of RNA-dependent polymerases. (see Fig. 1)

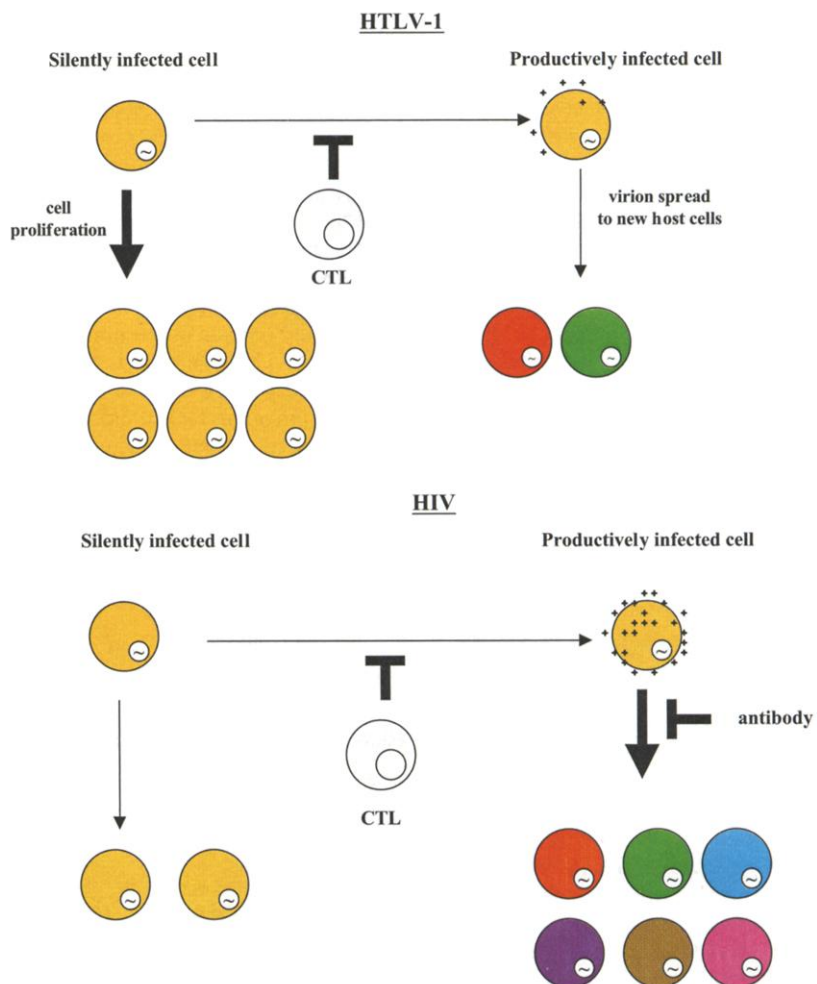
However, HTLV-1 is not transcriptionally silent *in vivo*, and a recent study has provided evidence that persistent virion replication contributes to the maintenance of the high proviral load of HTLV-1 (23). A high proportion of HTLV-1 provirus-positive T cells spontaneously express the viral transactivator protein Tax within a few hours of isolation *in vitro* (24). *In vivo*, the *tax* gene is subject to positive selection in healthy HTLV-1 carriers (18); this selection is probably exerted by the strong, chronically activated cytotoxic T lymphocyte (CTL) response to HTLV-1 (24, 25), which is mainly directed against a single viral antigen, the Tax protein (26). The CTL-mediated selection exerted on the *tax* gene can result in the appearance of putative escape mutations in CTL epitopes of Tax (27). However, there are two important factors that limit the impact of CTL-mediated selection on the rate of sequence variation and evolution of HTLV-1. First, the sequence variation within one infected host is constrained, because a cell that starts to express the Tax protein is likely to be killed by the abundant activated Tax-specific CTLs before it can complete the viral replication cycle (24, 25). Because most nucleotide changes in retroviruses arise during reverse transcription, the abortive replication cycles do not result in the accumulation of sequence variants. As mentioned above, the high load of HTLV-1 is maintained mainly by the proliferation of HTLV-1 provirus-carrying lymphocytes. Second, even if a sequence variant (such as a CTL escape mutant) does become established in the proviral gene pool in one individual, it is unlikely to carry a survival advantage during transmission, because the new hosts are likely to differ in their class I human leukocyte antigen genotype. Therefore, CTL-mediated selection does not accelerate evolution of HTLV-1 in the population.

Although two infected people are likely to differ in the T cell epitopes that they recognize, there is much less variation between individuals in the epitope specificity of the antibody response. The importance of the antibody response to HTLV-1 has not been fully established, but antibodies are in general less important than T cells in controlling the rate of replication of persistent noncytopathic viruses (28). Thus, although antibody-mediated selection is probably a major factor that drives the evolution of cytopathic viruses such as HIV and influenza A (29), it has less influence on the evolution of a persistent noncytopathic virus such as HTLV-1.

#### Lentiviruses: HIV-1 and SIV

HIV-1 is a complex lentivirus that evolved from a similar virus, SIV, found in nonhuman primates. Unlike HTLV and FeLV, HIV-1 causes a single disease syndrome, and it does so

in essentially all untreated, infected individuals. Given the uniform outcome of infection, it might appear surprising that HIV-1 exhibits a much higher degree of genetic diversity. However, HIV not only replicates at a high rate, but also is cytopathic and so causes considerable cell turnover. The rapid turnover of both the virus and infected cells allows the emergence of many mutations in the viral genome during both reverse transcription and transcription by RNA polymerase (Fig. 1). Nucleotide sequences from epidemiologically unlinked individuals from different parts of the world who are infected with the major HIV-1 group differ by 30% in parts of the genome such as the envelope gene. Remarkably, only about one-third of the nucleotide positions in the coding sequences of the HIV-1 genome are invariant among available full-length genomes (30), indicat-



**Fig. 1.** Immune selection and sequence variation in retrovirus infections. Retroviruses can replicate by two different routes. Mitotic division of a provirus-carrying cell replicates the viral genome faithfully and can occur without viral gene expression. This first "mitotic" route of replication therefore results in a relatively uniform population of virus sequences. The full cycle of virion production requires the action of the highly error-prone viral reverse transcriptase, and the viral gene expression exposes the viral antigens to attack by the host immune system. In this second "infectious" route of retrovirus spread, the frequent reverse transcriptase-generated mutations and immune-mediated selection can result in a population of viruses that is genetically highly diverse. Each color represents a cell infected with a different sequence variant of the retrovirus.

ing extraordinary genomic plasticity.

Within the first few months of HIV-1 infection, there is evidence for variants that have escaped cellular immune responses (31, 32). Unlike HTLV, CTL responses are directed at epitopes in several HIV-1 proteins. There is also strong selection for viruses that can evade humoral immune responses, and antibody escape variants have been shown to have increased replication fitness in the SIV-macaque model system (33). In people who progress more slowly to disease and/or have a low virus load, there is greater viral sequence diversity than in rapid progressors, as well as intense positive selection indicated by a high Dn/Ds ratio in the evolving viral sequences (34–36). This has led to the suggestion that diversity depends on the strength and duration of the host immune response. An alternative model to explain these associations is that a slow progressor is one who is infected with a less fit strain that replicates relatively poorly, and is therefore under strong positive selection in the host to increase its replication fitness. In the SIV model, the levels of virus replication are clearly dependent on the properties of the infecting isolate; viral loads and disease outcome are typically less variable among animals infected with the same virus than between groups of animals infected with different viruses (33). As yet there is no direct evidence that SIVs that have higher replication fitness are more genetically stable; such evidence would enable a test of the model that diversity and viral burden are both defined in some manner by the fitness of the infecting variant. It seems likely that both viral genetics/fitness and host selection pressures determine the extent of diversity and rate of progression.

As for FeLV, different target cells may be important for HIV-1 transmission versus HIV-1 persistence. At least two molecules are required for HIV entry into cells: the CD4 molecule and a multiple membrane-spanning chemokine receptor such as CCR5 or CXCR4. The viruses that are found within the first few months after HIV infection almost invariably require the CCR5 coreceptor for entry, which suggests that CCR5 variants are favored for transmission (37). In support of this model, it has been shown that individuals who do not express cell surface CCR5 as the result of a specific genetic polymorphism are less susceptible to HIV infection (38, 39). In a significant fraction of persons in the later stages of infection, viruses may either switch to use the CXCR4 coreceptor or adapt to use multiple coreceptors (37), thus providing for flexibility in the cells that can be infected in a chronically infected host. Interestingly, in some cases a mutation in HIV-1 that affects cell tropism also alters immune recognition, making it difficult to

assess which aspect of selection drives variation (40). SIV does not appear to switch coreceptors in the same manner, yet viruses that evolve during SIV infection replicate more efficiently in the host (33). One attractive model to explain this increased replication is that viruses may be selected if they are less dependent on high levels of cell surface expression of CD4 or CCR5 to infect cells productively. Selection forces that act on three retroviruses and the results of these forces are summarized in Table 1.

### Consequences and Complications of Retroviral Variation

Success in treating retroviral infections has required a better understanding of their enormous potential for genetic adaptation. For example, single-drug therapy against HIV-1 rapidly failed because it was simply too easy for the virus to generate a single mutation that would permit resistance. It was only when the virus was presented with multiple simultaneous puzzles (in the form of a combination of antivirals) that it could be held at bay for significant periods. Variation can also present major hurdles for vaccine development because it can lead to immune escape. Moreover, the choice of vaccine strains has been complicated by the fact that the viruses that are transmitted may be a subset of HIV-1 variants. It seems likely that vaccines that are based on the genome of a transmitted strain, and that provide the broadest array of common dominant epitopes, will have the best chance of providing some protection. Indeed, the only successful retroviral vaccine developed to date is the one against FeLV, and the most effective formulation of this vaccine uses the transmitted variant, FeLV-A, as immunogen (41).

The most serious consequence of the genetic flexibility of retroviruses is the effect that mutation and selection have on disease. The host can often coexist for many years with the

virus that was transmitted, and genetic variation (in the virus and in the host) determines both the probability and the course of disease. It has been shown in both the FeLV and SIV models that viruses that emerge in the later stages of retroviral infection are more pathogenic than the transmitted form of the virus (10, 33). However, although HTLV-1-associated diseases typically develop after a clinically silent period of many years, there is no evidence that specific variant HTLV-1 genomes are responsible. There is no apparent selection for viruses that can infect new target cells during HTLV-1 infection, perhaps because HTLV-1 is maintained chiefly by mitosis. In contrast, selection for new cell targets that results from continual rounds of virus replication appears integral to the disease process in other retroviral infections.

It has hitherto been impossible to reliably detect rare or transient mutations in retroviral infections. Therefore, much of our understanding of the selection pressures acting on retroviruses relies on the analysis of variants after they become dominant in the population. There is a need for methods that allow the detection and quantification of rare mutations, because this would permit detailed characterization of the effects of specific selection forces on the abundance of these mutant viral genomes in real time. These genetic data should be complemented with parallel, direct measurements of effector cell populations, including virus-specific CTL, T helper, and B cell clones, to identify which aspects of selection are most critical at different stages of infection and, in turn, how selection influences disease progression. For such data to be useful, they must be combined with mathematical and computational methods, which are increasingly useful in understanding the viral dynamics that underlie the disease process (42). This powerful combination of modeling and experiment is needed to understand, for example, the relative contributions of virus-

**Table 1.** Variation and selection in retroviruses.

	HTLV	FeLV	HIV	Effect on diversity
Source of mutations				
Replication through mitosis	++++	++	+	Low
Viral turnover → new rounds of reverse transcription	+	++	++++	High
Selection force				
CTL	+++	++*	+++	Variable†
Antibody	+/-	++*	+++	Variable
Cell tropism	—	+++	++	High
Extent of variation				
Within the host	Low	Moderate‡	High	
Between hosts	Low	Low	High	

\*Because little is known regarding FeLV immune selection, these are predictions. †The effect of immune selection on sequence diversity depends on the strength of the immune response and the degree to which the virus can be detected by the host. The latter is affected by the manner in which the virus amplification occurs (see Fig. 1). ‡This does not include variation that occurs through recombination with cellular sequences, which may lead to a nucleotide sequence difference of ~20% relative to the parental virus (6).

mediated toxicity and immune-mediated cytotoxicity to the rapid turnover of HIV-1-infected cells, and to understand the relative contributions of mitotic (proviral) and infectious (virion) spread of retroviruses within the host. This understanding may in turn directly influence drug treatment and vaccine strategies for retroviral infections.

# References and Notes

1. X. Wei *et al.*, *Nature* **373**, 117 (1995).
2. D. D. Ho *et al.*, *Nature* **373**, 123 (1995).
3. M. Nei, T. Gojobori, *Mol. Biol. Evol.* **3**, 418 (1986).
4. L. M. Mansky, *J. Virol.* **74**, 9525 (2000).
5. J. L. Rohn, J. Overbaugh, in *Persistent Viral Infections*, I. S. Y. Chen, R. Ahmed, Eds. (Wiley, Chichester, UK, 1999), pp. 379–408.
6. M. M. Anderson *et al.*, *Science* **287**, 1828 (2000).
7. O. Jarrett, P. H. Russell, *Int. J. Cancer* **21**, 466 (1978).
8. P. R. Donahue *et al.*, *J. Virol.* **62**, 722 (1988).
9. J. Overbaugh *et al.*, *Science* **239**, 906 (1988).
10. H. M. Temin, *Trends Genet.* **7**, 71 (1991).
11. J. Overbaugh *et al.*, *Nature* **332**, 731 (1988).
12. Y. Takeuchi *et al.*, *J. Virol.* **66**, 1219 (1992).
13. M. van Zeijl *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 1168 (1994).
14. D. G. Miller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 78 (1994).
15. M. P. Kavanaugh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7071 (1994).
16. F. L. Cosset *et al.*, *J. Virol.* **69**, 6314 (1995).
17. Y. Ina, T. Gojobori, *J. Mol. Evol.* **31**, 493 (1990).
18. S. Niewiesk *et al.*, *J. Virol.* **68**, 6778 (1994).
19. J. P. Slatery *et al.*, *Genome Res.* **9**, 525 (1999).
20. T. Uchiyama, *Annu. Rev. Immunol.* **15**, 15 (1997).
21. M. Nagai *et al.*, *J. Neurovirol.* **4**, 586 (1998).
22. M. Cavois *et al.*, *Oncogene* **12**, 2419 (1996).
23. G. P. Taylor *et al.*, *J. Virol.* **73**, 10289 (1999).
24. E. Hanon *et al.*, *Blood* **95**, 1386 (2000).
25. C. R. M. Bangham, *Curr. Opin. Immunol.* **12**, 397 (2000).
26. M. Kannagi *et al.*, *Int. Immunol.* **3**, 761 (1991).
27. S. Niewiesk *et al.*, *J. Virol.* **69**, 2649 (1995).
28. R. M. Zinkernagel, H. Hengartner, *Immunol. Today* **18**, 258 (1997).
29. R. M. Bush *et al.*, *Science* **286**, 1921 (1999).
30. B. Korber *et al.*, in *The Changing Face of HIV and AIDS*, R. Weiss *et al.*, Eds. (Oxford Univ. Press, Oxford, in press).
31. D. Price *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1890 (1997).
32. P. Borrow *et al.*, *Nature Med.* **3**, 205 (1997).
33. J. T. Kimata *et al.*, *Nature Med.* **5**, 535 (1999).
34. E. L. Delwart *et al.*, *J. Virol.* **68**, 6672 (1994).
35. S. M. Wolinsky *et al.*, *Science* **272**, 537 (1996).
36. V. Lukashov *et al.*, *J. Virol.* **69**, 6911 (1995).
37. T. L. Hoffman, R. W. Doms, *AIDS* **12**, S17 (1998).
38. R. Liu *et al.*, *Cell* **86**, 367 (1996).
39. M. Samson *et al.*, *Nature* **382**, 722 (1996).
40. A. McKnight *et al.*, *J. Virol.* **69**, 3167 (1995).
41. E. A. Hoover *et al.*, *J. Am. Vet. Med. Assoc.* **199**, 1392 (1991).
42. M. A. Nowak, R. M. May, *Virus Dynamics* (Oxford Univ. Press, New York, 2000).
43. J.O. is an Elizabeth Glaser Scientist. We thank M. Emerman, Q. Sattentau, J. Weber, and M. McClure for comments on the manuscript.

## VIEWPOINT

# Population Biology of Multihost Pathogens

Mark E. J. Woolhouse,\* Louise H. Taylor, Daniel T. Haydon

The majority of pathogens, including many of medical and veterinary importance, can infect more than one species of host. Population biology has yet to explain why perceived evolutionary advantages of pathogen specialization are, in practice, outweighed by those of generalization. Factors that predispose pathogens to generalism include high levels of genetic diversity and abundant opportunities for cross-species transmission, and the taxonomic distributions of generalists and specialists appear to reflect these factors. Generalism also has consequences for the evolution of virulence and for pathogen epidemiology, making both much less predictable. The evolutionary advantages and disadvantages of generalism are so finely balanced that even closely related pathogens can have very different host range sizes.

Most pathogens are capable of infecting more than one host species. This includes the 60% of human pathogen species that are zoonotic (1), causing diseases of major public health concern such as influenza, sleeping sickness, Lyme disease, food poisoning, and variant CJD. It also includes more than 80% of pathogens of domestic animals (2), notably those causing 57 of the 70 livestock diseases of greatest international importance (3), such as rinderpest, foot-and-mouth disease, and heartwater. Pathogens such as influenza A virus, rabies virus, and *Blastocystis hominis* can infect hosts not only of different species but from different orders or classes (2). Yet, despite their ubiquity and importance, multihost pathogens have been largely neglected by population biologists in favor of the simpler paradigm of a single-host species.

Many, though not all, pathogens that can infect multiple hosts can also be transmitted by

multiple hosts, and these can be regarded as ecological generalists rather than specialists. The advantages of generalism are poorly understood: it has been suggested that evolution should favor specialism, either because of the existence of functional trade-offs that limit the fitness of generalists in any one habitat or because evolution may proceed faster within narrower niches (4); these arguments apply especially to pathogens because they are under selection pressure to coevolve with their hosts (5). Yet paradoxically, only a minority of pathogens are specialists in the sense that they exploit a single host species.

So what processes lead to pathogens having multiple hosts, and why do multihost pathogens seem so pervasive? The evolution of generalism requires that pathogens have both the capability to exploit potential alternative host species and the opportunity to transmit to them. The subsequent maintenance of generalism depends on the consequences of an increased host range for pathogen population biology, especially such features as pathogenicity and epidemiology.

## Capability to Infect Multiple Hosts

Pathogens are usually, though not always, less infectious to a different host species. This is referred to as the species barrier (6), and there are two main strategies for overcoming it. Some pathogens have an inherent ability to infect multiple host species; for example, *Trypanosoma brucei rhodesiense* has a number of variant surface glycoprotein genes that encode for receptors with different affinities to specific mammalian transferrins (7). More commonly, pathogens produce many different genetic variants, some of which become associated with different host species, e.g., rabies (8). Gene products involved in host specificity have been identified for some pathogens, such as human immunodeficiency virus (HIV), mouse hepatitis virus, and *Citrobacter rodentium* (9).

Genetic change associated with host switching constitutes host adaptation. This may involve a small number of nucleotide substitutions or more major genetic changes such as reassortment, e.g., influenza A (10), or the acquisition of genetic elements (sometimes associated with virulence as well as host specificity), e.g., *Salmonella typhimurium* (11). Host adaptation can be so rapid that pathogen lineages adapt to different host tissues (12) or to vector versus host cells (13).

Species barriers are routinely crossed by some pathogens (such as rabies virus, which is regarded as a true multihost pathogen), but much more rarely by others [such as simian immunodeficiency virus, which is thought to have been transmitted to humans from other primates only very rarely and to have diverged rapidly into new single-host pathogens, HIV-1 and HIV-2 (14)]. Another example of pathogen

Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian EH25 9RG, UK.

\*To whom correspondence should be addressed. E-mail: mark.woolhouse@ed.ac.uk