Opinion TRENDS in Microbiology Vol.13 No.10 October 2005

The evolutionary dynamics of endogenous retroviruses

Aris Katzourakis, Andrew Rambaut and Oliver G. Pybus

Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK

Endogenous retroviruses (ERVs) are vertically transmitted intragenomic elements derived from integrated retroviruses. ERVs can proliferate within the genome of their host until they either acquire inactivating mutations or are lost by recombinational deletion. We present a model that unifies current knowledge of ERV biology into a single evolutionary framework. The model predicts the possible long-term outcomes of retroviral germline infection and can account for the variable patterns of observed ERV genetic diversity. We hope the model will provide a useful framework for understanding ERV evolution, enabling the testing of evolutionary hypotheses and the estimation of parameters governing ERV proliferation.

Endogenous retroviruses in vertebrate genomes

Retroviruses convert their RNA genome into DNA and integrate into the genome of their host in the form of a provirus. Occasionally, proviral integrations occur in germline cells and are therefore transmitted vertically [1]. Most vertebrates contain traces of past retroviral germline integrations, collectively called endogenous retroviruses (ERVs) [2]. After the initial integration, ERVs can copy themselves to different locations within the genome, which gives rise, over long periods of time, to a family of related ERV elements [3]. ERVs are classified amongst retroelements – mobile genetic elements that proliferate within the genomes of their hosts and that use an RNA intermediate during replication. Other retroelements include the human short and long interspersed nuclear elements (SINEs and LINEs).

The explosive growth in available genomic sequence data has enabled the study of ERV diversity and evolution. Using bioinformatics tools, >98 000 human ERVs (HERVs) have been identified, constituting ~5% of the genome [4,5]. These sequences are currently classified into 31 families (or lineages), each resulting from a distinct infection of the germline [6]. Complete phylogenies have been reconstructed for several HERV families; these vary in size, genetic variability and topology. Some HERV phylogenies are unusually 'star-like' in shape, with short internal branches and long external branches (Table 1; e.g. Figure 1a,c). The size and shape of HERV family phylogenies are a direct result of past evolutionary processes; therefore, we can use such phylogenies to reconstruct the dynamics of endogenous retroviruses in the human genome.

Here we outline a simple model of ERV evolutionary dynamics, which predicts both the eventual outcome of a germline infection and the expected shapes of ERV family phylogenies, under different evolutionary scenarios. This provides a framework for understanding and measuring the parameters that determine ERV proliferation and loss, and for testing hypotheses regarding ERV evolution. In these models, ERV proliferation and loss are considered on an evolutionary timescale, thus, each gain or loss event represents the effective fixation or removal of an element within the host population.

Mechanisms of ERV proliferation and loss

Most HERV families are thought not to be currently proliferating [4], with the probable exception of HERV-K(HML2) [7–9], although ERV families in other

Table 1. The variable size and shape of HERV family phylogenies[a]

HERV family	Number of HERV elements	Phylogeny imbalance (B_1 statistic)[b]	Phylogeny starlike-ness (γ statistic)[c]	Phylogeny depth (genetic distance from root to tips)[d]
HERV-E	34	15.9 (15.4, 16.5)*	−3.6 (−3.8, −3.4)*	0.144 (0.139, 0.150)
HERV-F(b)	23	10.7 (9, 12.2)	−7.1 (−7.2, −6.9)*	0.098 (0.094, 0.104)
HERV-K(HML5)	37	19.3 (18.5, 20.2)	−8.1 (−8.3, −8)*	0.122 (0.118, 0.125)
HERV-S	16	8.2 (6.8, 9)	−5.4 (−5.6, −5.3)*	0.142 (0.135, 0.150)
HERV-K(HML2)	44	16.7 (15.2, 18.4)*	−0.9 (−1.1, −0.7)	0.096 (0.091, 0.100)

[a]The B_1 statistic measures tree imbalance [49]. Imbalanced trees are 'comb-like', that is, the number of tips falling on either side of each branching point tend to be different (e.g. Figure 1e). The γ statistic measures tree starlike-ness [22]. Starlike trees have short internal branches and long terminal branches (e.g. Figure 1a,c). Asterisks indicate values that reject a simple null model of phylogeny shape (the constant-rate birth process) [22]. Upper and lower confidence intervals are shown in parentheses. Values were obtained from previously published HERV sequence alignments [7,11] using a Bayesian Markov Chain Monte Carlo (MCMC) approach, implemented in the computer program BEAST [50].
[b]The B_1 imbalance statistic depends on tree size and can be compared directly only if the number of elements are identical.
[c]Negative γ values are more starlike.
[d]Units are expected nucleotide substitutions per site.

Corresponding author: Pybus, O.G. (oliver.pybus@zoo.ox.ac.uk).
Available online 16 August 2005

 0966-842X/$ - see front matter doi:10.1016/j.tim.2005.08.004

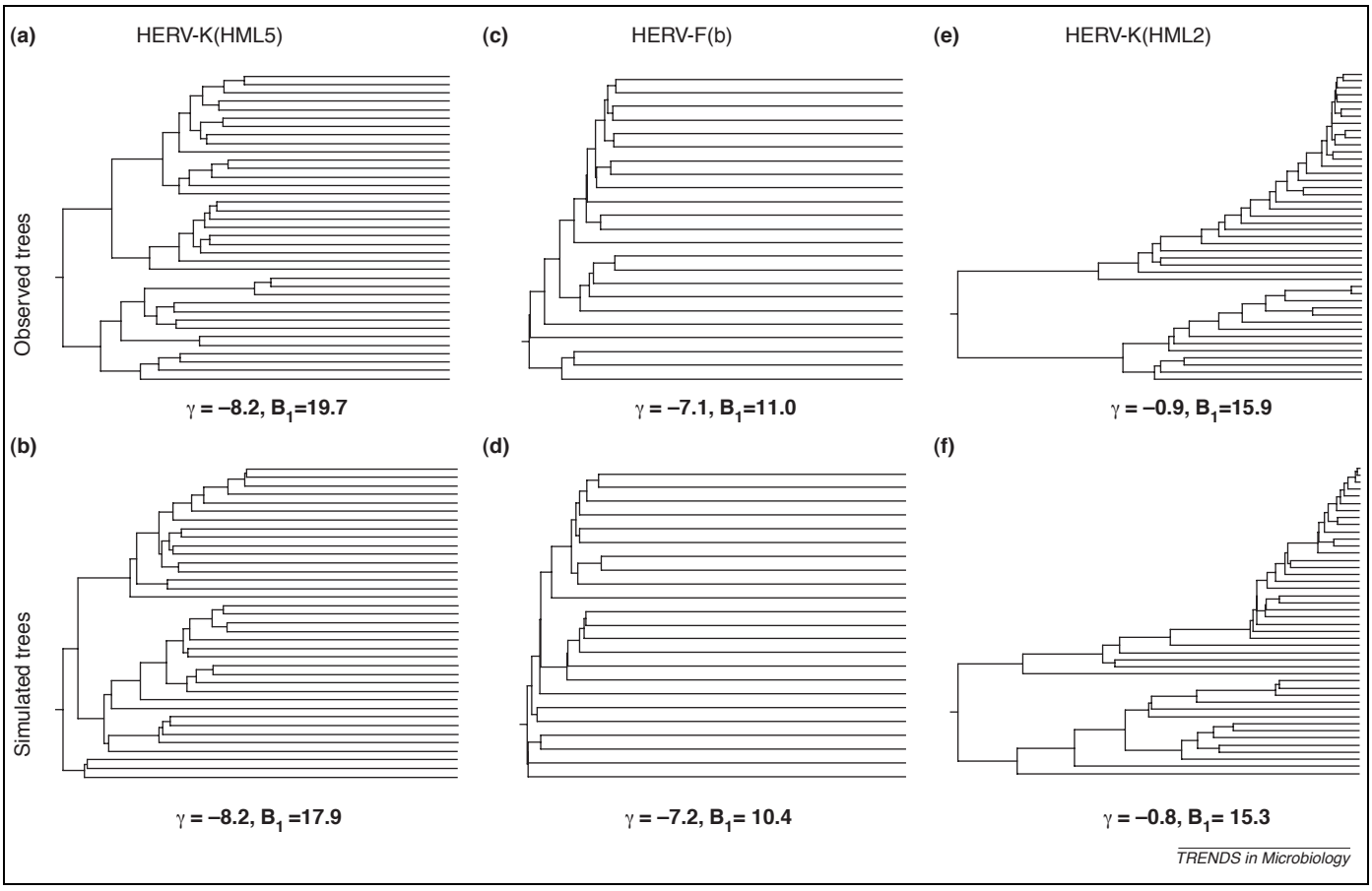

Figure 1. Comparison of observed and simulated HERV family phylogenies. **(a), (c)** and **(e)** are phylogenies of the HERV-K(HML5), HERV-F(b) and HERV-K(HML2) families, reconstructed from real data. **(b), (d)** and **(f)** are selected simulated phylogenies of the same size, obtained using the model in Box 2 (simulation software available from the authors on request). Each simulation was performed for a length of time directly proportional to the phylogeny depth in Table 1. Tree imbalance and starlike-ness statistics (B_1 and γ) are reported for each phylogeny. Phylogeny (b) was simulated using parameters $b=1.3$, $d=0.05$, $p=0.72$ and $i=0.35$; phylogeny (d) was simulated using parameters $b=6$, $d=0.05$, $p=0.95$ and $i=0.35$; phylogeny (f) was simulated using parameters $b=1.1$, $d=0.1$, $p=0.65$ and $i=0.35$. The units of *b, d* and *i* are per lineage, per unit genetic distance.

species are active today [10]. There are several mechanisms by which ERVs can proliferate and these are becoming increasingly well understood [2]. Proliferation can occur by an intra-cellular retrotransposition process within the germline, either in *cis* (where the virus supplies all the necessary proteins for replication) or in *trans* (where defective viruses are complemented by proteins from other viruses). Alternatively, ERVs might form infectious exogenous particles and thereby reinfect germline cells. It is possible to distinguish among these mechanisms by determining whether purifying selection has been acting on the retroviral genes *gag*, *pol* and *env*. Reinfection requires functional copies of all three genes, whereas retrotransposition in *cis* requires only *gag* and *pol*. Retrotransposition in *trans* requires an intact promoter, located within the long terminal repeats (LTRs), and other motifs for the expression and packaging of viral RNA, but does not require any functional retroviral genes. A comparison of purifying selection amongst genes shows that the majority of HERV families have proliferated by reinfection [7,11]; we concentrate on that mechanism here. The purifying selection observed in HERV families is consistent with multiple rounds of reinfection between proliferation events, as is the case with murine leukaemia virus [12,13]. A small number of HERV families appear to have retrotransposed and, interestingly, these have reached greater copy numbers in the genome than their reinfecting counterparts [11].

There is a good chance that new ERV elements will acquire mutations during reinfection or retrotransposition because both mechanisms involve an error-prone reverse transcriptase replication. Some mutations, such as stop codons and frameshift mutations, will render new elements immediately inactive (i.e. unable to proliferate by reinfection or retrotransposition in *cis*). By contrast, reactivating back mutations will be improbably rare. Once integrated into the genome, ERV elements will also accumulate host-induced inactivating mutations as a result of replication errors during host cell division. Additionally, recombination between the proviral LTRs will remove the coding region of the ERVs, leaving behind only a 'solo LTR' [14,15]. These can be thought of as 'fossil' ERVs that accrue cumulatively through time – hence ~85% of HERV elements are present in this deleted form [4,8].

Modelling the evolutionary dynamics of ERVs

Current models of retroelement proliferation are informal and fall into a continuum. At one extreme is the 'strict master gene' model [16], in which just one actively proliferating element exists and produces only inactive

progeny. In opposition, the 'transposon' [17] or 'random template' [16] model supposes that all elements are active and equally likely to produce progeny. Between these extremes lies the 'intermediate' model [18], which supposes that a proportion of retroelements in a family are capable of proliferating. Such models were suggested initially for SINE and LINE elements [17], and have been referred to subsequently in the ERV literature [19–21]. However, these models have not been formalized so quantitative comparison and testing with real data are not possible.

We introduce here a quantitative framework that explicitly incorporates mechanisms of ERV reinfection and inactivation, and which unifies the continuum of models outlined earlier. The framework can be implemented as: (i) a set of differential equations that describes the long-term deterministic dynamics of ERV proliferation (Box 1), or (ii) a stochastic branching process that describes the shapes of observed ERV phylogenies (Box 2). The former does not consider stochastic effects that are important when the number of ERV elements in a family is small. Both implementations consider how the number of active and inactive elements in a family change through time and depend on four parameters: b, the rate at which active elements replicate by reinfection; p, the proportion of progeny that are inactivated by mutation during reinfection; i, the rate at which elements are inactivated by mutation during host cell division; and d, the rate at which any element is removed by recombinational deletion (Box 1). The size of an ERV family will depend on the balance between the generation of new active elements (determined by b and p) and their loss or

Box 1. The dynamics of retrovirus reinfection

If reinfection is the dominant or sole route of ERV replication then the number of elements in an ERV family can be represented by a pair of differential equations, which characterize the number of replication-active (A) and replication-inactive (I) elements in the family.

$$\frac{dA}{dt} = bA(1-p) - iA - dA$$

$$\frac{dI}{dt} = bAp + iA - dI$$

Each active element proliferates by reinfection, such that rate b equals the rate of fixation of new elements in the host population. A proportion p of new elements are replication-inactive, as a result of mutations accumulated during reinfection. Host germline cell division might also generate inactivating mutations – this occurs at rate i per active element. Finally, both active and inactive elements can be removed from the genome, at rate d, by recombinational deletion. As with rate b, rate d represents the fixation of the deletion event in the host population.

The behaviour of this model is surprisingly simple and has three outcomes:

(i) If $bp > (i+d)$ then both A and I will end up increasing indefinitely at an exponential rate. Under some parameter values, I will temporarily decrease to begin with.

(ii) If $bp < (i+d)$ then the opposite occurs and both A and I will ultimately decrease towards zero. However, depending on the parameters used, I might temporarily increase initially.

(iii) If $bp = (i+d)$ then A will remain unchanged at its initial value, A_0, and I will tend towards an equilibrium value of $A_0 d^{-1}(b\text{-}bp\text{-}i)$.

Outcome (i) leads to runaway ERV proliferation that will ultimately lead to destruction of the genome of the hosts unless ERV replication is controlled in some way. Outcome (ii) can lead to ERV families that initially grow in size but which are fated from the outset to decline at a rate dependent on d and inevitably disappear from the genome of the hosts. Outcome (iii) represents an unstable equilibrium and will not occur in nature.

Box 2. A stochastic branching process model

A more realistic model of ERV evolution can be obtained by converting the deterministic model (Box 1) into a stochastic branching process, which represents the number of active and inactive elements at time t as A_t and I_t, respectively. The stochastic and deterministic models behave similarly when A_t and I_t are large. The stochastic process considers each short time interval, Δt, during which each active element can undergo the following events:

(*i*) Production of a new active element by reinfection, with probability $b(1\text{-}p)\,\Delta t$.

(*ii*) Production of a new inactive element by reinfection, with probability $bp\Delta t$.

(*iii*) Inactivation by mutation during host cell division, with probability $i\Delta t$.

(*iv*) Recombinational deletion, with probability $d\Delta t$.

Similarly, in each short time interval Δt, each inactive element undergoes:

(*v*) Recombinational deletion, with probability $d\Delta t$.

This process can generate phylogenetic trees that represent the evolutionary history of ERV family proliferation (see Figure I). Each branch in the phylogeny corresponds to an ERV element; new phylogenetic lineages are generated by reinfection (event types *i* and *ii*) and are removed by deletion (event types *iv* and *v*). Inactivating host mutations (event type *iii*) convert an active lineage into an inactive one. Each inactive element corresponds to a single external phylogeny branch (or just part of one), whereas each active element corresponds to one or more internal phylogeny branches (and perhaps part of an external branch too; see Figure I).

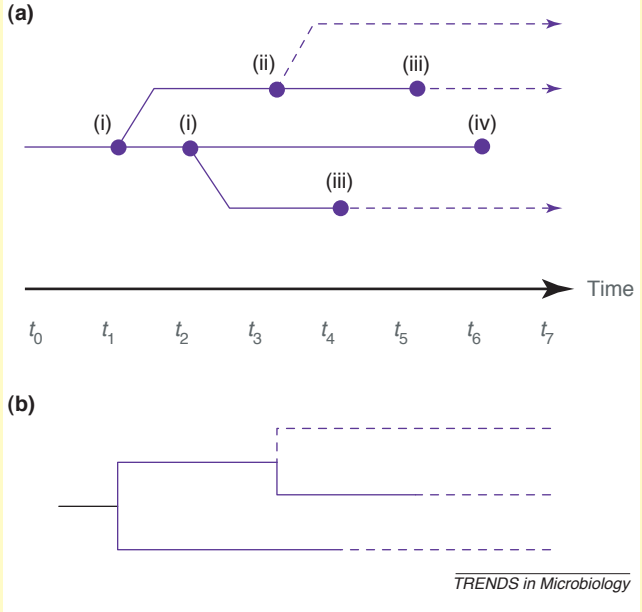

Figure I. The stochastic branching process. Solid and dashed lines represent active and inactive elements, respectively. Filled circles represent evolutionary events, occurring at times t_0, t_1, t_2, etc. Numbers in parentheses correspond to the event definitions in the box text. **(a)** The history of an ERV family, beginning at time t_0 with one active element. Reinfection produces two new active elements at t_1 and t_2, and a new inactive element at t_3. Two active elements are inactivated by host mutations at time t_4 and t_5. The remaining active element is deleted at time t_6. The host genome is sampled at t_7, when three inactive elements are observed. **(b)** An ERV family phylogeny, reconstructed from the host genome sampled at time t_7. The phylogeny is generated from (a) by removing the deleted elements.

inactivation (determined by i and d). The complete inactivation of the family will result in its eventual extinction by recombinational deletion.

The deterministic model can be used to predict the long-term evolutionary outcome of ERV proliferation (Box 1). Depending on relative rates of reinfection, inactivation and deletion, ERV lineages in nature have just two fates. Either the total number of elements in a family (active and inactive) will indefinitely increase, thereby ultimately destroying the host genome, or the opposite will occur and the family will be eventually removed from the genome (Box 1). Under the former outcome, uncontrolled proliferation would provide a strong selective pressure for the evolution of host defensive mechanisms (see subsequent sections). Under the latter outcome, active elements are progressively lost, so proliferation eventually ceases. However, this outcome still enables an ERV family to increase in size for a period of time before going extinct (depending on the parameters used). Thus, it is possible that some of the sizeable ERV families that we observe are in fact evolutionary 'dead ends' that are pre-destined for extinction.

Stochastic effects will have a major role in ERV evolutionary dynamics when the numbers of active or inactive elements are small. These effects can be accounted for by converting the model in Box 1 into a stochastic branching process, such that the parameters b, i and d represent per-lineage event probabilities (Box 2). Similar models are often used in macro-evolutionary studies to investigate long-term trends in speciation and extinction (eg. [22]). The stochastic model can generate simulated ERV phylogenies to be compared with those obtained from real data. Figure 1 provides a proof-of-principle illustration of this approach; phylogenies simulated under selected parameter values can match the variable shapes of observed ERV trees.

Most importantly, the branching process can reproduce the asymmetric and strongly star-like nature of some ERV phylogenies. Taken at face value, such trees suggest that the rate of ERV proliferation has changed dramatically at some point in time. It would be reasonable to ascribe a biological cause to this change, such as the evolution of host defensive mechanisms (see next section). However, our model demonstrates that the observation of star-like phylogenies alone is not sufficient to warrant this conclusion because such trees can be generated by a simple 'null model', in which all parameters remain constant through time. The apparent discontinuity arises from the loss of the last active element, after which no further proliferation occurs.

Host defensive mechanisms

In future it should be possible to develop tests of the null model described earlier. Rejection of this model for particular ERV lineages would suggest changing reinfection, inactivation or deletion rates through time. A likely cause of such change is the evolution of host defensive mechanisms to prevent runaway ERV amplification (Box 1). Defensive mechanisms can target different stages of the retroviral replication cycle. For example, retrovirus cell entry can be blocked by mutations in host-cell surface receptor genes or by the expression of factors that block these receptors. Post-entry mechanisms include host-encoded inhibitors that interact with the retroviral capsid protein before genome integration, called post-entry restriction factors [23]. For example, the mouse *Fv1* gene inhibits infection by murine leukaemia virus [24]. Recent studies suggest that similar restriction factors exist in primates, including humans [25–28]. Intriguingly, *Fv1* is derived from a *gag* gene that belongs to a mouse ERV family (MuERV-L), a close relative of the human HERV-L family [24,29,30]. Other restriction factors might be retroviral in origin [31], although primate restriction factors discovered so far are attributed to the action of the TRIM5α protein [32–34] and do not appear to involve ERV proteins. It is tempting to speculate that ancient ERV elements have been co-opted by their hosts to restrict the ERV family from which co-opted elements are derived, or related ERV families [35]. The apparent lack of ERV-based restriction factors in humans [4] could be explained if such factors became redundant millions of years ago and have since accumulated neutral mutations.

Other ERV-limiting innovations include methylation, which can reduce the activity of transposable elements by transcriptional silencing or by promoting mutation at methylated sites [36]. Both processes appear to have some role in HERV-K(HML2) family activity [37]. Another source of potentially lethal mutations are the host-encoded APOBEC enzymes. In particular, APOBEC3G has been shown recently to induce hypermutation in *Vif*-defective HIV but it also appears to have been under diversifying selection throughout the last $\sim$30 million years of primate evolution [38]. HIV and other lentiviruses arrived in primates much more recently (perhaps $<$1 million years ago [39]), therefore, it seems likely that longer-standing infectious agents, such as ERVs, might have driven this selection [38,40]. Experimental evidence has shown that APOBEC3G can induce hypermutation on murine intracisternal A-particles (IAP) and MusD elements [41] but the range of the ERV-limiting ability of APOBEC3G across viral lineages and host taxa is unknown.

ERVs could cause disease by many mechanisms, including deleterious effects arising from integration into, or near to, functional genes, thereby disrupting or altering gene expression. Intriguingly, new theories of cancer suggest that the fitness cost of such events might depend on the life history of the hosts. It is proposed that larger, longer-lived animals have evolved mechanisms to control cancer [42] because such animals present more opportunities for cancer to develop before reproduction. By analogy we might also expect these animals to be more prone to ERV-related diseases before reproduction. The theory provides two explanations for the greater number of active ERV families in mice [43] than in humans [7–9]. First, if new active elements in humans are more likely to cause disease before reproduction then they are less likely to be fixed in the population. Second, if the fitness cost of ERV integration in mice is lower then selection for host defence mechanisms should be weaker in that species and more elements will be observed.

Future directions

A key goal for future research is to implement the model presented here in a statistical inference framework, which would enable hypothesis testing and estimation of model parameters. This could be achieved by statistically comparing the output of the stochastic model (Box 2) with real data, perhaps using approximate Bayesian summary statistic methods [44]. Estimation of model parameters should improve our understanding of ERV dynamics (although, potentially some parameters could be strongly correlated and thus difficult to estimate independently). For example, the potential of HERVs to cause disease means that it is important to investigate the current activity of HERV lineages. This can be addressed by searching for insertion polymorphisms within and among populations, as has been done for the HERV-K(HML2) family [9,14,45–47]. However, a combination of genomic data and new statistical tools could help to assess the current activity of ERV families, even if insertion polymorphism data are unavailable. In future, the models outlined here could be extended to include the genomic location of ERV elements, which could affect both expression levels and the probability of recombinational deletion [48]. In addition, the models could perhaps be adapted to investigate the evolutionary dynamics of other genomic elements, such as retrotransposons.

The completion of genome sequencing projects in other species, such as the mouse [43] and chimpanzee, will enable a comparative genomics approach to ERV evolutionary history. ERV evolutionary behaviour might vary among species as a result of many factors, including stochastic forces, retroviral genetic differences and variation in host biology. Distinguishing amongst these factors will be a fundamental issue in future ERV research.

Acknowledgements

We thank Paul Harvey for useful comments on this article. A.K. is funded by an MRC fellowship, and A.R. and O.G.P. are funded by the Royal Society.

References

1 Vogt, P.K. (1997) Historical introduction to the general properties of retroviruses. In *Retroviruses* (Coffin, J.M. *et al*., eds), pp. 1–25, Cold Spring Harbor Laboratory Press
2 Boeke, J.D. and Stoye, J.P. (1997) Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In *Retroviruses* (Coffin, J.M. *et al*., eds), pp. 343–435, Cold Spring Harbor Laboratory Press
3 Tristem, M. (2000) Identification and characterisation of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* 74, 3715–3730
4 Lander, E.S. *et al*. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
5 Paces, J. *et al*. (2002) HERVd: database of human endogenous retroviruses. *Nucleic Acids Res.* 30, 205–206
6 Katzourakis, A. and Tristem, M. (2005) Phylogeny of human endogenous and exogenous retroviruses. In *Retroviruses and Primate Genome Evolution* (Sverdlov, E.D., ed.), pp. 186–203, Landes Bioscience
7 Belshaw, R. *et al*. (2004) Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4894–4899
8 Stoye, J.P. (2001) Endogenous retroviruses: Still active after all these years? *Curr. Biol.* 11, R914–R916
9 Turner, G. *et al*. (2001) Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* 11, 1531–1535
10 Gifford, R. and Tristem, M. (2003) The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26, 291–316
11 Belshaw, R. *et al*. (2005) High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol. Biol. Evol.* 22, 814–817
12 Lock, L.F. *et al*. (1988) Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice. *EMBO J.* 7, 4169–4177
13 Tomonaga, K. and Coffin, J.M. (1999) Structures of endogenous nonecotropic murine leukemia virus (MLV) long terminal repeats in wild mice: Implication for evolution of MLVs. *J. Virol.* 73, 4327–4340
14 Hughes, J.F. and Coffin, J.M. (2004) Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: Implications for human and viral evolution. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1668–1672
15 Seperack, P.K. *et al*. (1988) Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *Proc. Natl. Acad. Sci. U. S. A.* 85, 189–192
16 Clough, J.E. *et al*. (1996) Computer simulation of transposable element evolution: Random template and strict master models. *J. Mol. Evol.* 42, 52–58
17 Deininger, P.L. *et al*. (1992) Master genes in mammalian repetitive DNA amplification. *Trends Genet.* 8, 307–311
18 Cordaux, R. *et al*. (2004) Retrotransposition of Alu elements: how many sources? *Trends Genet.* 20, 464–467
19 Buzdin, A. *et al*. (2003) Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. *Genomics* 81, 149–156
20 Costas, J. (2001) Evolutionary dynamics of the human endogenous retrovirus family HERV-K inferred from full-length proviral genomes. *J. Mol. Evol.* 53, 237–243
21 Medstrand, P. and Mager, D.L. (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72, 9782–9787
22 Pybus, O.G. and Harvey, P.H. (2000) Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. Biol. Sci.* 267, 2267–2272
23 Bieniasz, P.D. (2003) Restriction factors: a defense against retroviral infection. *Trends Microbiol.* 11, 286–291
24 Best, S. *et al*. (1996) Positional cloning of the mouse retrovirus restriction gene *Fv1*. *Nature* 382, 826–829
25 Besnier, C. *et al*. (2002) Restriction of lentivirus in monkeys. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11920–11925
26 Cowan, S. *et al*. (2002) Cellular inhibitors with *Fv1*-like activity restrict human and simian immunodeficiency virus tropism. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11914–11919
27 Hatziioannou, T. *et al*. (2003) Restriction of multiple divergent retroviruses by Lv1 and Ref1. *EMBO J.* 22, 385–394
28 Towers, G. *et al*. (2000) A conserved mechanism of retrovirus restriction in mammals. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12295–12299
29 Bénit, L. *et al*. (1997) Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a *gag* coding sequence closely related to the *Fv1* restriction gene. *J. Virol.* 71, 5652–5657
30 Qi, C.F. *et al*. (1998) Molecular phylogeny of *Fv1*. *Mamm. Genome* 9, 1049–1055
31 Stoye, J.P. (2002) An intracellular block to primate lentivirus replication. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11549–11551
32 Keckesova, Z. *et al*. (2004) The human and African green monkey TRIM5α genes encode Ref1 and Lv1 retroviral restriction factor activities. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10780–10785
33 Yap, M.W. *et al*. (2004) TRIM5α protein restricts both HIV-1 and murine leukemia virus. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10786–10791
34 Hatziioannou, T. *et al*. (2004) Retrovirus resistance factors Ref1 and Lv1 are species-specific variants of TRIM5α. *Proc. Natl. Acad. Sci. U. S. A.* 101, 10774–10779
35 Best, S. *et al*. (1997) Endogenous retroviruses and the evolution of resistance to retroviral infection. *Trends Microbiol.* 5, 313–318
36 Yoder, J.A. *et al*. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340
37 Lavie, L. *et al*. (2005) CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J. Virol.* 79, 876–883

38 Sawyer, S.L. *et al*. (2004) Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2, E275
39 Sharp, P.M. *et al*. (1999) Origins and evolution of AIDS viruses. *Biol. Bull.* 196, 338–342
40 Holmes, E.C. (2004) Adaptation and immunity. *PLoS Biol.* 2, 1267–1269
41 Esnault, C. *et al*. (2005) APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433, 430–433
42 Leroi, A.M. *et al*. (2003) Cancer selection. *Nat. Rev. Cancer* 3, 226–231
43 Waterston, R.H. *et al*. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
44 Beaumont, M.A. *et al*. (2002) Approximate Bayesian computation in population genetics. *Genetics* 162, 2025–2035
45 Macfarlane, C. and Simmonds, P. (2004) Allelic variation of HERV-K(HML-2) endogenous retroviral elements in human populations. *J. Mol. Evol.* 59, 642–656
46 Burmeister, T. *et al*. (2004) Insertional polymorphisms of endogenous HERV-K113 and HERV-K115 retroviruses in breast cancer patients and age-matched controls. *AIDS Res. Hum. Retroviruses* 20, 1223–1229
47 Mamedov, I. *et al*. (2004) A rare event of insertion polymorphism of a HERV-K LTR in the human genome. *Genomics* 84, 596–599
48 McVean, G.A.T. *et al*. (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* 304, 581–584
49 Kirkpatrick, M. and Slatkin, M. (1993) Searching for evolutionary patterns in the shape of a phylogenetic tree. *Evolution Int. J. Org. Evolution* 47, 1171–1181
50 Drummond, A.J. and Rambaut, A. (2003) BEAST v1.0, Available from http://evolve.zoo.ox.ac.uk/beast/