The human zoo: endogenous retroviruses in the human genome

Lindell Bromham

The main focus of the human genome sequencing project has been gene discovery, but a great additional benefit is that it offers the chance to examine the large proportion of the genome that does not contain human genes. The nature of this 'noncoding' DNA is poorly understood, both as an evolutionary question (how did it get there?) and in the functional sense (what is it doing now?). Much of the noncoding DNA is derived from retroviruses that have inserted their DNA into the genome. The availability of complete genomic sequences will revolutionize studies of the number and location of endogenous retroviruses, their role in genome evolution, and their contribution to human disease.

The human genome contains surprisingly few genes [1,2]; most of the DNA in our genome apparently does not encode human proteins or RNA. Genome sequencing projects promise breakthroughs in understanding the origins and maintenance of this 'noncoding' DNA, much of which is made up of repeat elements. Some are simple repeats, such as the between two-to-five nucleotide repeats that make up microsatellites, which may result from template slippage during replication [3]. Others are more complex, and are at least partially responsible for their own replication and dispersal throughout the genome. For example, Class II transposable elements use transposase, encoded in their own DNA or borrowed from another source, to excise their DNA from one location and re-insert it elsewhere in the genome [4]. Here, I focus on endogenous retroviruses (ERVs) and the structurally similar long terminal repeat (LTR) retrotransposons, which are Class I transposable elements that replicate by retrotransposition (Box 1).

All viruses replicate by hijacking the cellular resources of their host. Retroviruses multiply by using their own enzyme, reverse transcriptase, to make a DNA copy of their RNA genome, which is then inserted into the host genome where it is transcribed by the transcription machinery of the host. If a retrovirus is inserted into germline DNA it will be copied and inherited with the rest of the host genome as an ERV. A surprisingly large proportion of the human genome is ultimately derived from viral genomes in this way [1].

What is this viral DNA doing in our genome? There are four possible fates of an ERV residing in our DNA: (1) it retains the ability to produce virus particles; (2) it retains transpositional activity; (3) virus sequences are co-opted into host genome function; and (4) virus sequences decay into 'junk' (noncoding) DNA. These four fates, which are not mutually exclusive, leave a patchwork of viral DNA in the human genome, with a variety of roles and effects. In addition, a fifth fate – excision from the host DNA – might either entirely delete the retrovirus or leave a partial retroviral sequence, such as a solo LTR. The first fate of an ERV leaves a store of 'live' viruses embedded in the genome, and all fates have important implications for host genome function and evolution. The availability of complete human genome sequences will revolutionize the study of the role of ERVs in the evolution and function of the human genome.

Lindell Bromham
Biological Sciences, University of Sussex, Falmer, Brighton, UK BN1 9QG.
e-mail: L.D.Bromham@sussex.ac.uk

How many ERVs are there in the human genome?

The systematic sequencing of all the DNA in the genome makes it possible to locate and describe all retroelements in the genome, including those not actively expressed and those with incomplete sequences. Analyses of the recently released draft human genome sequences suggest that at least 8% of the genome is derived from LTR retroelements, with perhaps as many as half a billion copies [1,2]. Availability of genomic sequence data has already led to the discovery and characterization of new ERVs, and to re-examination of the systematics and classification of human endogenous retroviruses (HERVs) [5–7]. Tristem [6] has described 22 HERV families, using a molecular phylogeny of retroviral sequences to select monophyletic groups with characteristic primer binding sites. One such family, HERV-K, contains many human-specific proviruses with virtually intact genomes [8], including at least one HERV that is functionally complete [1] and capable of producing virus particles [9].

As well as cataloguing the presence of ERVs, databases of genomic DNA (and mRNA) sequences can be used to investigate the origins and activity of retroelements. ERVs arise by two processes: entry into the genome from another individual (by infection), and replication within the genome (by transposition). Both processes influence genome evolution and function through an increase in retroelement copy number, shuffling of host DNA, and the disruption of normal gene function by insertion of retroelements. Furthermore, both infection and transposition could impact upon human health, either directly (through viral replication or expression of viral products) or indirectly (through transpositional mutagenesis – insertion in or near host genes).

http://tree.trends.com 0169-5347/02/$ – see front matter PII: S0169-5347(01)02394-1

Box 1. Retroelements

Retroelements are a broad class of genetic sequences that arise through retrotransposition: DNA is copied from an RNA template by reverse transcriptase and inserted into the host genomic DNA [a]. This review is concerned primarily with endogenous retroviruses (ERVs) and long terminal repeat (LTR) retrotransposons, collectively referred to as LTR retroelements. These share the common feature of being flanked by LTRs that encode promoters and enhancers necessary for retrotransposition. The major difference between the types of LTR retroelements is the presence or absence of genes that code for proteins associated with reproduction and virulence (Fig. I).

ERVs are copied from a viral RNA genome, and inserted into the host genome as proviruses. ERVs that retain all the functional sequences of the viral genome might be capable of producing infectious particles [e.g. some human endogenous retrovirus K (HERV-K) retroelements], but many others have mutations that render the coding sequences defective, or the LTRs nonfunctional [e.g. endogenous retrovirus 9 (ERV-9)]. LTR retrotransposons retain transposition activity by having functional LTRs, but lack some or all of the coding sequences needed to produce infectious particles. Some LTR retroelements produce their own polymerase to promote transposition, whereas others borrow the polymerase produced by another retroelement. Mouse VL30 retroelements, for example, have functional LTRs with strong promoters but generally do not code for functional proteins, although they show some sequence homology to murine leukemia virus (MLV)-like ERVs. VL30 retroelements have packaging sequences that allow them to co-package with MLV virions and thus cross-infect cells [b]. Poly(A) retrotransposons resemble ERVs but have only a single 5′ copy of the LTR, with a poly(A) tail at the 3′ end; they are nonetheless capable of autonomous retrotransposition, often encoding proteins necessary for transposition. For example, long interspersed nuclear elements (LINEs) actively transpose in both mice and humans, and are responsible for relatively high rates of insertional mutagenesis [c].

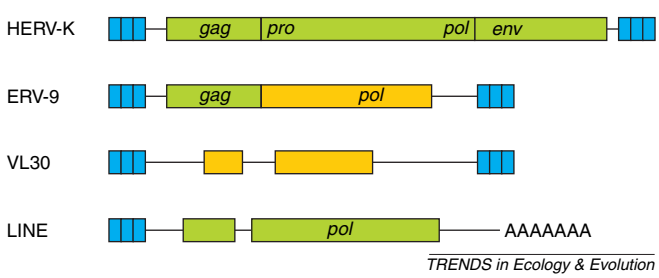

Fig. I. Retroelements vary in structure. Intact retroviruses have long terminal repeats (LTRs) and coding sequences [e.g. human endogenous retrovirus K (HERV-K)], but retroelements can retain transposition activity with incomplete coding sequences [e.g. endogenous retrovirus 9 (ERV-9)], with no functional genes (e.g. the mouse retroelement VL30) or without paired LTRs [long interspersed nuclear element [LINE)]. Key: Blue, LTR; Green, intact gene; Yellow, defective coding sequence.

References

a 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. 343–436, Cold Spring Harbor Laboratory Press
b French, N.S. and Norton, J.D. (1997) Structure and functional properties of mouse VL30 retrotransposons. *BBA – Gene Struct. Expr.* 1352, 33–47
c Kazazian, H.H. (1999) An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* 22, 130–131

Where do ERVs come from?

The first fate of an ERV is to retain the ability to replicate as a virus, capable of infecting other cells and, potentially, other individuals. The existence of 'live' viruses embedded in mammalian genomes raises fears of cross-infection. In particular, the ability of ERVs to move between species has caused doubts over the safety of xenotransplantation (use of animals as organ donors for humans), prompting the search for active retroviruses in donor species, and examination of their activity and infectivity (Box 2).

Phylogenies constructed from DNA sequences can reveal cross-species origins of ERVs. For example, molecular phylogenies suggest that ERVs have entered the human genome on many different occasions, from a wide range of free-living animal virus groups [6,7]. More generally, molecular phylogenies can demonstrate the propensity of ERVs to jump between species. ERVs being passively copied with the host genome should have a phylogeny matching that of their host, whereas discordant trees indicate host switching [10]. Martin and co-workers [11] surveyed ERVs related to the murine leukemia viruses from a range of vertebrate species, and constructed a molecular phylogeny based on a DNA sequence from the polymerase gene. The phylogeny of the ERVs from birds and reptiles closely reflected the phylogeny of their hosts, indicating that most of these ERVs were passed on only through the host germline. The mammalian ERVs, however, showed a very different pattern: the ERV phylogeny bore little relation to the host phylogeny, indicating a high level of host switching for these sequences (Fig. 1). Specific cases of cross-species infection might prove informative for investigating the mechanisms of transfer of ERVs from their host genome to another, often distantly related, genome. One example is the intriguing relationship between the koala retrovirus (KoRV) and the 'gibbon-ape leukemia virus' (GALV) (Box 3).

How do ERVs jump between species?

Molecular phylogenies for HERVs and other mammalian ERVs indicate that ERVs occasionally move between species. Are humans at risk of novel ERV infections? Do current agricultural practices and medical techniques exacerbate this risk? To answer these questions we need not only to catalogue the occurrence and activity of ERVs in humans and other species [12] but also to understand the factors that prompt an ERV to emerge as an exogenous virus, capable of infecting other species.

It has been suggested that the activity of transposable elements in the genome might be

Box 2. ERVs and xenotransplantation

The increasing shortfall between the number of available donor organs and the number of patients requiring organ transplants has led to the development of xenotransplantation – the use of other species as organ donors for humans [a]. Using animals specially bred and raised for the purpose reduces the risk of transmission of animal viruses and bacteria, but ERVs are unavoidably present in the genome of each cell of a xenotransplanted organ. The risk of releasing an ERV infection via xenotransplantation is not insignificant, given that mammalian ERVs seem capable of cross-species jumps (as demonstrated by both evolutionary studies and laboratory experiments). Furthermore, ERVs that are often benign in their usual host can be virulent when they cross into a new one [b].

The animal most commonly targeted as a xenotransplant donor is the domestic pig. Pig genomes contain many porcine endogenous retroviruses (PERVs) [c], and pig cells in culture can produce PERV virus particles that are capable of cross-infecting human cells [d]. The same can be expected of other potential donor species; baboons, for example, carry BaEV, a replication-competent endogenous retrovirus that is capable of infecting human cells [e]. Although there is, as yet, no evidence of PERV activity introduced by xenotransplanted organs [f,g], the spectre of releasing novel ERVs into the human population has prompted calls for a moratorium on xenotransplantation until the risks of ERV emergence are better understood [h].

References

a Takeuchi, Y. and Weiss, R.A. (2000) Xenotransplantation: reappraising the risk of retroviral zoonosis. *Curr. Opin. Immunol.* 12, 504–507

b 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. 343–436, Cold Spring Harbor Laboratory Press

c Patience, C. *et al.* (2001) Multiple groups of novel retroviral genomes in pigs and related species. *J. Virol.* 75, 2771–2775

d Patience, C. *et al.* (1997) Infection of human cells by an endogenous retrovirus of pigs. *Nat. Med.* 3, 282–286

e Allan, J.S. (1998) The risk of using baboons as transplant donors – exogenous and endogenous viruses. *Ann. New York Acad. Sci.* 862, 87–99

f Herring, C. *et al.* (2001) Monitoring xenotransplant recipients for infection by PERV. *Clin. Biochem.* 34, 23–27

g Switzer, W.M. *et al.* (2001) Lack of cross-species transmission of porcine endogenous retrovirus infection to nonhuman primate recipients of porcine cells, tissues, or organs. *Transplantation* 71, 959–965

h Furst, I. (1999) Swiss ban xenotransplants. *Nat. Biotechnol.* 17, 316

triggered by external factors or by changes in genome regulation [13]. Barbara McClintock, pioneering researcher in the field of transposable elements, suggested that transposition could be triggered by environmental stress or by hybrid dysgenesis (genetic incompatibility between parent genomes in hybrids) [14]. Furthermore, it is often implied that a stress-moderated increase in transposition could be considered a host-mediated mechanism for generating genetic variants, by rearrangement of host DNA and the alteration of gene expression patterns [15]. This hypothesis views retroelements as a symbiotic part of the host genome, and transpositional mutagenesis as effectively something the genome is doing to itself, to 'remodel' itself as a rapid response to environmental challenge, or as a means of generating genetic change at speciation to speed the process of adaptation to new environments.

An alternative interpretation is that increased transposition activity reflects loss of genome control, such that replication of transposable elements runs unchecked, potentially at the expense of the host [16]. This view of retroelements as 'selfish' elements, parasitic on the host genome, is more plausible considering that it is unlikely that disruption to an organized genome will result in beneficial change. Instead, it will usually have negative consequences (such as interruption of normal gene expression) or none at all (if inserting into nonfunctional DNA). Furthermore, several observations suggest that retroelements are actively removed or suppressed by the host genome, rather than being maintained as useful symbionts. The high copy number of single HERV LTRs has been considered to be evidence that HERVs are actively excised from the human genome [17].

DNA methylation also appears to play a role in regulating transposon activity and expression [18,19]. For example, reduction in methylation in a hybrid between two wallaby species apparently led to massive amplification of a kangaroo endogenous retrovirus (KERV-1), greatly expanding chromosomes around the centromeres [13]. Similarly, HERV activity is often higher in cancer cells (which might have undergone changes in methylation or chromatin structure) than in normal cells [20]. The evolution of resistance to infection by ERVs, for example by mutation of cellular receptors that allow viral entry into the cell or through the development of immune-like responses to infection, also suggests that the genome is trying to limit ERV activity [21].

An increase in transposition activity through loss of host genome control might also explain why ERVs appear to move more readily between cell lines than between intact organisms. For example, pig cells in culture can produce porcine endogenous retrovirus (PERV) particles, which can infect human cells [22], but there has been no detectable expression of PERVs in humans (or other primates) with xenotransplanted pig tissue [23,24]. However, PERVs can infect immunodeficient mice that have received pig cells as xenotransplants [25,26]. Genome regulation might be compromised in cultured cells such that they are more likely to shed ERVs, and that they lack the appropriate immune-like response to ERVs, thus making them vulnerable to infection (Box 3). These observations provide reason to suspect that risk of zoonoses of ERVs could be increased by the use of animal cell cultures in medical research, and possibly also by the increasing number of patients with compromised immune systems, such as organ recipients or people with AIDS

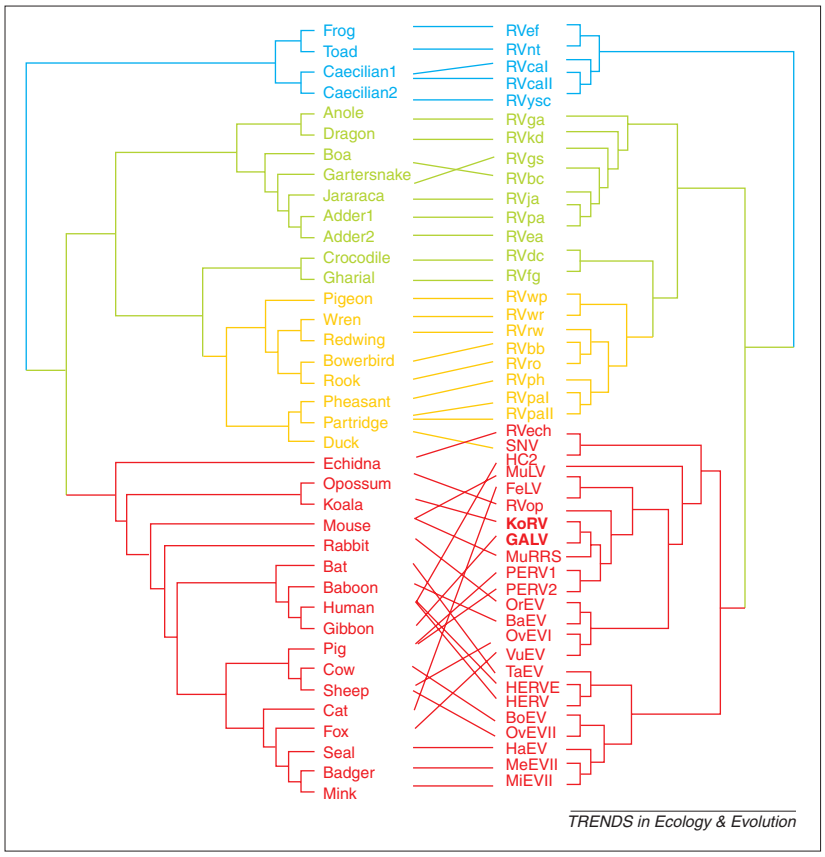

Fig. 1. Endogenous retroviruses (ERVs) can jump species. The phylogeny of many ERVs, for example some ERVs of amphibians (blue), reptiles (green) and birds (yellow) matches that of their hosts. But the phylogeny of some other ERVs, as exemplified by this molecular phylogeny of a sample of mammalian ERVs (red) based on part of the polymerase gene, does not closely follow the host phylogenetic pattern, indicating a high degree of cross-species transmission in these cases. Although the phylogenetic relationships are uncertain, and the exact degree of host switching in mammalian ERVs is dependent on the accuracy of the phylogenies, it is clear that the virus tree is not concordant with the host tree. Redrawn and adapted, with permission, from Ref. [11]. Taxon names have been abbreviated.

ERV transposition: multiplying in the genome

The second fate of an ERV is to retain the ability to replicate in the genome by transposition. Transposition could be an important factor in genome evolution and function, not only through increasing the copy number of retroelements but also through the incidental rearrangement of host DNA and the effects of retroelement insertion on gene expression and function. Understanding the frequency of transposition of retroelements in the human genome is important for assessing the role of ERVs in human disease. The availability of the complete human genome sequence will allow an assessment of the level of transposition by revealing transposed copies of retroelements, and by allowing estimation of the time frame of retroelement transposition. There have been several approaches towards estimating the frequency of transposition in the human germline over evolutionary time.

Frequency of HERV transposition

Determination of the insertion sites of ERVs within the genome allows a comparative approach to dating retroelement origins. Assuming that insertion is random, so that insertion of the same retrovirus at the same locus in different species is unlikely, ERV insertions shared across species reflect the presence of the retroelement in their last common ancestor. This approach has demonstrated that ERVs have entered the genome from the earliest periods of mammalian evolution right up to the recent evolutionary past [6,7,17,27]; the identification of human-specific ERVs suggests that new viruses have entered the human genome within the past five million years (Myr) of human evolution [8,28]. A study of ERV-L sequences (a family of ERVs present in all mammalian lineages) suggests that ERVs can remain capable of transposition even when they have been present in the genome for tens of millions of years [17].

Although retroelements have been implicated in human mutagenesis, largely because of the relatively high rate of long interspersed nuclear element (LINE) transposition [29] (Box 1), the failure to identify polymorphisms for HERV integrations across human populations suggests that HERV transposition is not frequent [7]. Whole genome sequences for many individuals would be ideal for identifying novel retroelement insertions. Both of the human genome sequences currently available [1,2] are amalgams of the DNA extracted from several individuals: although single-nucleotide polymorphisms between individuals are described, neither project reports positional polymorphisms for larger DNA sequences. Genomic sequences from individuals suffering from illnesses that have been linked (by some studies) to ERV expression – such as multiple sclerosis, schizophrenia and diabetes – might prove particularly informative in this respect. The potential for retroelement transposition (particularly early in embryonic development) to influence genetic expression could result in identical twins that vary in expression of genetic disease as a result of HERV-induced mutations [30,31], complicating the study of environmental and genetic factors in human diseases. It is therefore important that the frequency and impact of retroelement transposition is accurately estimated.

Dating transposition activity with molecular clocks

Molecular clock analyses, which use differences between DNA sequences to estimate the time since their last common ancestor, have been applied to dating retroelement activity in the human genome. Two different approaches have been taken. The International Human Genome Sequencing Consortium [1] compared DNA sequences across ERV families to estimate the number of substitutions in the proviral genomes since they entered the human genome. They reconstructed an ancestral sequence for each family, then estimated the number of substitutions that had occurred between the ancestral sequence and the observed retroelement sequences. They concluded that there has been virtually no LTR retroelement activity in the human genome for >25 Myr, and that most

Box 3. What is this koala virus doing in my gibbon?

One of the most fascinating examples of cross-species transmission of an endogenous retrovirus (ERV) is the close relationship between the koala retrovirus (KoRV) and the 'gibbon-ape leukemia virus' (GALV). KoRV is an endogenous virus of koalas which is associated with lymphoma, ubiquitous in koala populations [a]. GALV is an exogenous virus associated with leukemia-like illness in captive gibbons [a]. This unusual scenario raises interesting questions, the answers to which could have important consequences for epidemiology in both wildlife and humans. How does an endogenous virus from one species become an infectious disease of another? How could a virus jump between species whose natural distributions are separated by several thousand kilometres including the Arafura Sea?

The possibility of an intermediate vector that spans both species distributions is currently under investigation. An alternative explanation is laboratory-mediated transfer: because KoRV is widely distributed in both captive and wild koala populations but GALV is known only from captive gibbons associated with research facilities, it is possible that the marsupial virus was introduced to gibbons through a research-related activity, such as animal contact or the use of cell lines and their products. Indeed, a variant of this virus, GALVX, is known only from human cell lines [b]. Clarifying the relationship between KoRV and GALV could therefore be an important case study of the risk of cross-species ERV infection via medical research.

It would be interesting to survey for GALV in wild gibbons to determine whether GALV occurs naturally in gibbons or only in captive colonies. An accurate molecular date of the origin of GALV and GALVX would also shed light on the cross-species transmission of KoRV/GALV. However, it is difficult to calibrate a molecular clock for a virus that shows both endogenous (host genome) and exogenous (viral) replication [a]. Calibrating the molecular clock with a rate appropriate for an endogenous virus – such as the host neutral substitution rate – would suggest that KoRV and GALV split tens of millions of years ago, whereas assuming that these viruses evolve at a rate typical of exogenous retroviruses would suggest a more recent split, only decades ago. Clearly much more work needs to be done before we have a clear picture of how a koala virus came to infect gibbons, or vice versa.

References

a Hanger, J.J. *et al.* (2000) The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus (KoRV): a novel type-C retrovirus related to gibbon ape leukemia virus (GALV). *J. Virol.* 74, 4264–4272

b Burtonboy, G. *et al.* (1993) Isolation of a C-type retrovirus from an HIV-infected cell-line. *Arch. Virol.* 130, 289–300

retroelements are as old as the mammalian radiation. These estimates led to the dramatic claim that 'LTR-retroposons appear to be teetering on the brink of extinction' [1].

This finding is in conflict with estimates of retroelement activity that use a phylogenetic approach or a molecular clock based on LTRs, both of which imply retroelement activity during recent human evolutionary history. There are several problems with this molecular dating approach, such as the use of inferred ancestral sequences [32]. The most serious problem is shared with most molecular clock studies: the difficulty of establishing an appropriate calibration rate [33]. The substitution rate of ERVs could vary dramatically, depending both on the activity of the viruses and on the nature of selection on ERV sequences (Box 3). For example, an ERV undergoing active replication and transposition will be copied by error-prone reverse transcriptase, whereas a provirus dormant in the genome will be copied with much higher fidelity by the host replication machinery. Active ERVs might be under strong negative selection to maintain the integrity of the coding sequence, or, conversely, under positive selection to evade host immune response. Nonreplicating ERVs might evolve at the neutral substitution rate of the host genome, or be under selection to maintain the function of co-opted viral sequences.

This leads to a dilemma. If it is assumed that ERVs are inactive, the appropriate calibration would be drawn from the relatively slow human substitution rates [1,6], leading to the conclusion that the large number of changes observed between proviruses results from a long evolutionary period with little recent retroviral activity. But if ERVs are assumed to be active, the exceedingly high rates of substitution in retroviruses and transposable elements [34,35] are more appropriate calibrations, leading to the conclusion that retroelements are much younger than was previously thought and that the genome is subject to ongoing transposition activity.

These difficulties are partly overcome by an alternative approach to dating transposition activity that compares the flanking LTR sequences of each retroelement [6,27]. The LTR sequences that flank the proviral genome are identical (or near identical) upon insertion, because they are copied from a single template sequence in the viral genome. Once a retroelement has inserted in the host genome, it is copied by the host replication machinery and so acquires mutations at the same rate as the host DNA. These mutations might destroy the function of viral sequences, but, given that loss of function in a retroelement is unlikely to disadvantage the host, these mutations cannot be selected against. Therefore the LTR sequences of a provirus inserted in the human genome should acquire substitutions at the human neutral substitution rate. Because each LTR sequence will acquire mutations independently, the number of differences between the LTRs can be used as an indication of time since insertion. This method has produced dates of insertion compatible with a phylogenetic approach to dating, although it is not without problems, such as the confounding effect of gene conversion [27]. Care must still be taken in selecting an appropriate calibration rate (for example, the substitution rate in retroelements is influenced by the region of the genome they reside in [1]), and

methods of dating transposition events that present molecular dates with appropriate confidence limits should be developed [36].

The role of ERVs in genome building

Retroelements influence the evolution and function of the genome in a variety of ways. The mutagenic effects of retroelement transposition can be substantial [29], and include disrupting the function of a host gene by inserting into it, positioning viral promoters near host genes, and moving host DNA around the genome as a by-product of the transposition. Inserting into a gene is likely to knock out its function, which would frequently be deleterious. Because retroelement LTRs contain strong promoters and enhancers, ERVs that insert near a host gene can cause dysregulation of gene expression. The association between some ERVs and cancer (e.g. leukemia viruses) might be due to runaway expression prompted by the presence of retroviral promoters [9]. ERVs also prompt gene duplication [37], either directly, by providing a mechanism for duplicating blocks of DNA [38], or indirectly, through the cellular presence of ERV-derived reverse transcriptase. Reverse transcription could increase the incidence of reinsertion of mRNA transcripts, which become processed pseudogenes that might either decay or possibly evolve new functions. In general the effects of ERV-induced mutagenesis are expected to be negative, as for any means of mutating an otherwise organized genome. However, ERVs can contribute adaptively to host genome evolution by providing sequences that can be utilized by the host [18]. This is the third fate of an ERV: for part or all of its genome to be co-opted into host genome function.

Co-option of ERV sequences

Sequences derived from retroviruses perform a variety of useful roles in the host genome. Sometimes these involve intact retroelements, for example the function of an ERV as telomerase in *Drosophila*, and the proposed immunological function of some ERVs [21,39]. More common is the adoption of specific viral sequences. For example, a gene derived from an HERV is involved in the formation of the human placenta [40], and the presence of retroviral proteins might confer some degree of immunity against related exogenous viruses [9]. The co-option of strong retroviral promoters and enhancers appears to be even more common [41].

Retroviruses must successfully compete with the host for the use of its transcription machinery, and consequently they can have extremely effective promoters and enhancers. Many mammalian genes have adopted these efficient viral promoters [41]; for example, retroviral promoters determine the tissue specificity of human salivary amylase [42]. Analysis of mouse *Mus musculus* endogenous retrovirus (MmERV) has suggested there might even be mechanisms that enhance co-option of viral promoters into the host genome [5]. The MmERV polyadenylation signal resides in a repeat sequence that is tandemly duplicated in viral transcripts, and also provides the active polyadenylation signal for nonviral (host) mRNAs. Several features of this poly(A) element, such as flanking repeats and a palindromic sequence, might enhance duplication of the element and possibly its transposition across the genome, aiding its co-option into the expression of both host and viral genes. Is this a transposable element within a transposable element, reminiscent of Swift's recursive 'a flea hath smaller fleas that on him prey; and these have smaller fleas to bite 'em, and so proceed *ad infinitum*'? Analysis of genome sequences might reveal the existence of such elements (which could be called 'ad-ins', signifying not only the *ad infinitum* nature but also the add-in functionality they contribute to both viral and host sequences).

Decaying into junk

Genomes can therefore contain intact ERVs, potentially capable of producing virus particles, and partial ERVs, either as active retroelements or as sequences co-opted into host genome function. The fourth fate of an ERV is to be rendered nonfunctional by mutation and subsequently to decay into 'junk' DNA. Mutations occurring in the LTRs can inhibit transposition, and random changes to the coding sequences are likely to result in nonfunctioning products, such as proteins truncated by stop codons. Most HERVs are nonfunctional in this sense [6,19]. The accumulation of ERV DNA might contribute to the 'C-value paradox' [43] – the observation that much of the DNA in the genome is not associated with any obvious function in the host's development, and that the amount of DNA in a species' genome is not obviously related to developmental complexity. However, it remains unclear just how much of a burden retroelement DNA is to its host.

Retroelements can rapidly replicate in the genome even when there is a fitness cost to the host [18], behaviour consistent with the interpretation of transposable elements as 'selfish' DNA or genomic parasites [43,44]. Retroelement copy number is likely to have some effect on host fitness, if not from the disruption caused by transposition then from the energetic cost of the sheer bulk of nonhost DNA that must be replicated. The only experimental evidence, from *Drosophila*, suggests that there can be a measurable fitness effect of copy number of transposable elements [45]. However, the persistence of individuals or lineages having massively amplified retroelement burdens with little apparent ill-effect [13], and the huge variation between even closely related lineages in the amount of 'junk' DNA derived from retroelements, show that there is no simple answer to calculating the cost of retroelements to the host.

Acknowledgements
I thank Jeff McKee and Francis Clark for ideas, debate and encouragement, two anonymous referees for their detailed comments, and Pete Lockhart for organizing the DOOM '01 meeting.

Whose DNA is it anyway?

Are retroelements genomic parasites or beneficial symbionts needed for genome maintenance and evolution? There is no doubt that retroelements have played an important role in the evolution of our genome, with adaptive functions such as providing novel genes and strong promoters and shuffling host DNA. However, ascribing genome-building functions to retroelements is not equivalent to saying that ERVs are a necessary part of the mammalian genome. In spite of the occasional win, the genome pays a high price for the presence of retroelements, from disease to mutation to the burden of copying someone else's DNA.

Genome sequencing projects will bring light to the mystery of why so much of our DNA does not contain genes needed to build a human being. Our genome is made up of sequences inherited not only from our hominid ancestors but also from the retroviruses and other transposable elements that inhabit our DNA. Much of the DNA we refer to as 'noncoding' codes for these genomic passengers. The sequencing projects will give us the first complete view of this 'human zoo'. The noncoding DNA, once regarded as junk, is where the most interesting discoveries will be made.

References

1 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
2 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
3 Stephan, W. and Walsh, B. (2001) Repetitive DNA: evolution. In *Encyclopedia of Life Sciences*, Macmillan (http://www.els.net)
4 Braam, L.A.M. and Reznikoff, W.S. (2001) DNA transposition: classes and mechanisms. In *Encyclopedia of Life Sciences*, Macmillan (http://www.els.net)
5 Bromham, L. *et al.* (2001) Discovery of a novel murine type C retrovirus by datamining. *J. Virol.* 75, 3053–3057
6 Tristem, M. (2000) Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the Human Genome Mapping Project database. *J. Virol.* 74, 3715–3730
7 Bock, M. and Stoye, J.P. (2000) Endogenous retroviruses and the human germline. *Curr. Opin. Genet. Dev.* 10, 651–655
8 Barbulescu, M. *et al.* (1999) Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans. *Curr. Biol.* 9, 861–868
9 Lower, R. *et al.* (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* 93, 5177–5184
10 Page, R.D.M. (1994) Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Syst. Biol.* 43, 58–77
11 Martin, J. *et al.* (1999) Interclass transmission and phyletic host tracking in murine leukemia virus-related retroviruses. *J. Virol.* 73, 2442–2449
12 Patience, C. *et al.* (2001) Multiple groups of novel retroviral genomes in pigs and related species. *J. Virol.* 75, 2771–2775
13 Waugh O'Neill, R.J. *et al.* (1998) Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature* 393, 68–72
14 McClintock, B. (1983) The significance of responses of the genome to challenge (Nobel Prize Lecture). Reprinted in *The Dynamic Genome: Barbara McClintock's Ideas in the Century of Genetics* (Fedoroff, N. and Botstein, D., eds), pp. 174–192, Cold Spring Harbor Laboratory Press
15 Palevitz, B.A. (2000) Genetic parasites and a whole lot more: transposable elements generate DNA mutations, alter gene expression, and otherwise fuel genetic diversity. *Scientist* 14, 13
16 Hurst, G.D.D. and Werren, J.H. (2001) The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* 2, 597–606
17 Benit, L. *et al.* (1999) ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* 73, 3301–3308
18 Kidwell, M.G. and Lisch, D.R. (2001) Transposable elements, parasitic DNA and genome evolution. *Evolution* 55, 1–24
19 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. 343–436, Cold Spring Harbor Laboratory Press
20 Mager, D.L. (1999) Human endogenous retroviruses and pathogenicity: genomic considerations. *Trends Microbiol.* 7, 431
21 Best, S. *et al.* (1997) Endogenous retroviruses and the evolution of resistance to retroviral infection. *Trends Microbiol.* 5, 313–318
22 Specke, V. *et al.* (2001) Productive infection of human primary cells and cell lines with porcine endogenous retroviruses. *Virology* 285, 177–180
23 Herring, C. *et al.* (2001) Monitoring xenotransplant recipients for infection by PERV. *Clin. Biochem.* 34, 23–27
24 Switzer, W.M. *et al.* (2001) Lack of cross-species transmission of porcine endogenous retrovirus infection to nonhuman primate recipients of porcine cells, tissues, or organs. *Transplantation* 71, 959–965
25 Deng, Y.M. *et al.* (2000) Transmission of porcine endogenous retroviruses in severe combined immunodeficient mice xenotransplanted with fetal porcine pancreatic cells. *Transplantation* 70, 1010–1016
26 van der Laan, L.J.W. *et al.* (2000) Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice. *Nature* 407, 90–94
27 Johnson, W.E. and Coffin, J.M. (1999) Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. U. S. A.* 96, 10254–10260
28 Medstrand, P. and Mager, D.L. (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* 72, 9782–9787
29 Kazazian, H.H. (1999) An estimated frequency of endogenous insertional mutations in humans. *Nat. Genet.* 22, 130–131
30 Deb-Rinker, P. *et al.* (1999) Molecular characterization of a MSRV-like sequence identified by RDA from monozygotic twin pairs discordant for schizophrenia. *Genomics* 61, 133–144
31 Karlsson, H. *et al.* (2001) Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4634–4639
32 Lynch, M. (2001) The molecular natural history of the human genome. *Trends Ecol. Evol.* 16, 420–422
33 Bromham, L.D. *et al.* (1999) Growing up with dinosaurs: molecular dates and the mammalian radiation. *Trends Ecol. Evol.* 14, 113–118
34 Drake, J.W. (1999) The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. In *Molecular Strategies in Biological Evolution* (Vol. 870) (Caporale, L.H., ed.), pp. 100–107, New York Academy of Sciences
35 Mules, E.H. (1999) Fidelity of retrotransposon replication. In *Molecular Strategies in Biological Evolution* (Vol. 870) (Caporale, L.H., ed.), pp. 108–118, New York Academy of Sciences
36 Rambaut, A. and Bromham, L. (1998) Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* 15, 442–448
37 Fedoroff, N.V. (1999) Transposable elements as a molecular evolutionary force. In *Molecular Strategies in Biological Evolution* (Vol. 870) (Caporale, L.H., ed.), pp. 251–264, New York Academy of Sciences
38 Kulski, J.K. *et al.* (1999) Coevolution of PERB11 (MIC) and HLA Class I genes with HERV-16 and retroelements by extended genomic duplication. *J. Mol. Evol.* 49, 84–97
39 Villareal, L.P. (1997) On viruses, sex and motherhood. *J. Virol.* 71, 859–865
40 Mi, S. *et al.* (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789
41 Britten, R.J. (1997) Mobile elements inserted in the distant past have taken on important functions. *Gene* 205, 177–182
42 Ting, C.N. *et al.* (1992) Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* 6, 1457–1465
43 Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature* 284, 604–607
44 Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotypic paradigm and genome evolution. *Nature* 284, 601–603
45 Charlesworth, B. and Langley, C.H. (1989) The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* 23, 251–287