# Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans

Madalina Barbulescu*, Geoffrey Turner*, Michael I. Seaman†, Amos S. Deinard‡§, Kenneth K. Kidd† and Jack Lenz*

Background: Endogenous retroviruses contribute to the evolution of the host genome and can be associated with disease. Human endogenous retrovirus K (HERV-K) is related to the mouse mammary tumor virus and is present in the genomes of humans, apes and cercopithecoids (Old World monkeys). It is unknown how long ago in primate evolution the full-length HERV-K proviruses that are in the human genome today were formed.

Results: Ten full-length HERV-K proviruses were cloned from the human genome. Using provirus-specific probes, eight of the ten were found to be present in a genetically diverse set of humans but not in other extant hominoids. Intact preintegration sites for each of these eight proviruses were present in the apes. A ninth provirus was detected in the human, chimpanzee, bonobo and gorilla genomes, but not in the orang-utan genome. The tenth was found only in humans, chimpanzees and bonobos. Complete sequencing of six of the human-specific proviruses showed that full-length open reading frames for the retroviral protein precursors Gag–Pro–Pol or Env were each present in multiple proviruses.

Conclusions: At least eight full-length HERV-K genomes that are in the human germline today integrated after humans diverged from chimpanzees. All of the viral open reading frames and *cis*-acting sequences necessary for HERV-K replication must have been intact during the recent time when these proviruses formed. Multiple full-length open reading frames for all HERV-K proteins are present in the human genome today.

Addresses: *Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA. †Department of Genetics, ‡Department of Anthropology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520-8005, USA.

§Present address: School of Veterinary Medicine, University of California Davis, Davis, California 95616, USA.

Correspondence: Jack Lenz
E-mail: lenz@aecom.yu.edu

Received: 11 June 1999
Revised: 8 July 1999
Accepted: 8 July 1999

Published: 3 August 1999

Current Biology 1999, 9:861–868
http://biomednet.com/elecref/0960982200900861

 ISSN 0960-9822

## Background

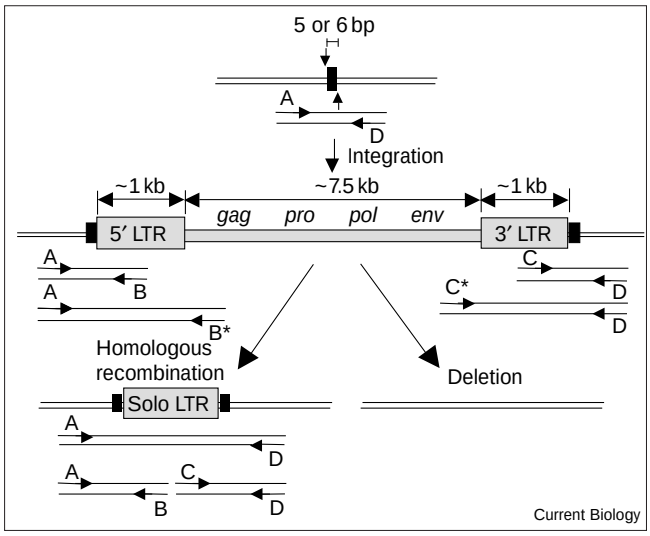
Endogenous retroviruses exist as proviruses (the integrated form of retroviral DNA) in the germ-line DNA of the host, which is transmitted from parent to offspring [1,2]. Formation of new germ-line proviruses can alter the structure and/or expression of cellular genes [3,4]. If endogenous retroviruses can replicate, they have the potential to cause disease [1], but endogenous retroviruses can also protect the host against retrovirally-induced diseases [5–8].

HERV-K was originally detected by low stringency hybridization with probes for the mouse mammary tumor virus and mouse intracisternal A particle [9,10]. It is one of several families of endogenous retroviruses in humans [1]. The human genome has been estimated to contain about 30–50 HERV-K proviruses [10]. DNA hybridization studies have shown that HERV-K proviruses are present in the genomes of humans, apes and cercopithecoids (Old World monkeys) [11–13]. More distantly related sequences have also been detected in platyrrhines (New World monkeys [14]. It is unknown whether the HERV-K proviruses that are in the human genome today formed long ago during evolution of the catarrhines (hominids, apes, and cercopithecoids) or whether these viruses continued to re-enter the human genome in recent times. We therefore cloned many of the full-length HERV-K proviruses in the human genome, determined when each formed and examined the structures of their genomes.

## Results

Individual retroviral proviruses have unique identities based on their positions within the host genome. Proviruses form at essentially random positions within the cellular genome [15]. The likelihood that two independent integrations will occur at the same site in the cellular genome is negligible [16]. Most of a provirus can disappear from a genome by homologous recombination between the two long terminal repeats (LTRs) to generate a solo LTR. Alternatively, part or all of a provirus may be removed by deletions extending into flanking cellular sequences or by gene conversion in cells hemizygous for the provirus (Figure 1). If the sequence of the cellular DNA flanking both sides of a provirus can be determined, PCR primers (Figure 1; primers A and D) can be generated that, when used in conjunction with primers in the viral genome, uniquely identify that provirus. Moreover, having primers that flank both sides of individual proviruses allows the preintegration site, provirus and solo LTR to be distinguished (Figure 1).

**Figure 1**




Strategy for the detection of proviruses. The double lines at the top show a preintegration site. Black box, sequence that is duplicated upon provirus formation; vertical arrows, the staggered cuts made by viral integrase. The double lines in the middle show the arrangement of genomic DNA after formation of the provirus. Sizes of portions of the HERV-K genome are indicated. The double lines at the bottom show two types of products that might derive from a proviral locus, a solo LTR and a deletion of a provirus that included flanking cellular sequences extending beyond the A and D primers. A, B, B*, C*, C and D represent PCR primers used to generate the indicated products.

To obtain clones of full-length HERV-K proviruses with flanking cellular sequences on both sides, human genomic libraries based on bacterial or phage P1 artificial chromosomes (BAC and PAC libraries, respectively) were screened. The large inserts in these clones compared with the 9.5 kb HERV-K genome made it likely that full-length proviruses would be recovered. Clones were isolated using a hybridization probe from the HERV-K *pol* gene; 90 clones were identified in a screen of BAC inserts of complexity equivalent to about 1.8 haploid human genomes. These clones were tested with a set of 11 PCR primer pairs that were spaced throughout the viral genome according to the sequence of HERV-K10 [17]; see Supplementary material for primer sequences. Of the 90 BAC clones, 25 were positive for most of the PCR markers. Virus–cell junction sequences were obtained for 11 of the clones; one provirus was found to be represented twice in the set of 11. The 10 proviruses identified in this manner were sequentially designated HERV-K101 through HERV-K110 (HERV-K101 and HERV-K102 are different from the two placental cDNA clones with the same names [14]). On the basis of identity of the flanking sequences, HERV-K107 and HERV-K110 are HERV-K10 and HERV-K18, respectively, identified by Ono *et al.* [10,17]. HERV-K108 is HERV-K(HLM-2.HOM) [18].

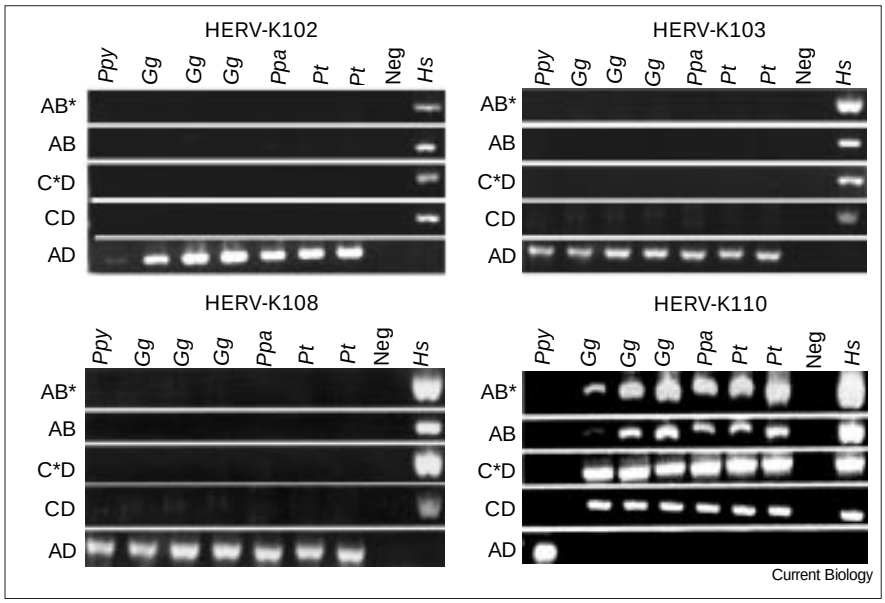
To determine when each of the 10 proviruses formed during human evolution, we performed PCR using, as template, genomic DNA from humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), gorillas (*Gorilla gorilla*) and an orang-utan (*Pongo pygmaeus*). Products for four of the proviruses are shown as examples (Figure 2). Eight of the proviruses were detected only in humans (Table 1). The distribution of each of the eight proviruses in genetically diverse humans was assessed by PCR using a set of DNAs from 35 humans. DNAs were tested from five individuals in each of seven different human populations (Biaka, Mbuti, Druze, Chinese, Nasioi, Mayan, and European), two of which were from sub-Saharan Africa. The eight proviruses were all present in all individuals of these populations, and no preintegration site alleles were detected (data not shown). Thus, all eight proviruses appear to have formed before the migration of modern humans across the world.

To test whether the lack of detection of the proviruses in apes was due to absence of the proviruses rather than false-negative PCRs, we determined whether the apes contained the preintegration sites. PCR primer pairs that flank both sides of each of the human-specific proviruses (Figure 1, primers A and D) were used to amplify the preintegration sites. The primer pairs used did not yield products from human DNA templates, except for those used for HERV-K107/HERV-K10. This showed that the primer pairs did not recognize non-allelic loci of any family of repeated sequences in the human genome. All of the ape genome templates yielded a product of the size expected for the preintegration sites for the human-specific HERV-K proviruses (Figure 2; Table 1).

Sequencing confirmed that the products amplified from the ape genomes were derived from authentic preintegration sites. When retroviruses integrate into a host genome, a short sequence of cellular DNA is duplicated at either end of the viral genome (Figure 1) [15]. Most HERV-K proviruses were flanked by 6 bp duplications of cellular DNA (Figure 3). HERV-K107/HERV-K10 had a duplication of only 5 bp. The ape sequences were precisely colinear with the human sequence except that they contained no proviral sequences and had only a single copy of the duplicated sequences that flank the human proviruses (Figure 3). Some single nucleotide differences were also observed. We conclude that all of the 10 proviruses except HERV-K105 and HERV-K110 formed after humans diverged from chimpanzees. The alternative explanation that any of the proviruses formed before the divergence, and that a proviral allele persisted at low frequency in apes is unlikely. It is highly improbable that such a polymorphism would have been retained at any of the loci in an ape species for the necessary length of time unless selective pressure maintained both alleles throughout the period [19].

**Figure 2**

PCR products for four of the proviral loci from humans and apes. The primer pairs used for PCR are indicated at the left of each gel photograph. A, B, B*, C*, C and D correspond to the primers whose priming sites are shown in Figure 1. *Ppy*, *P. pygmaeus*; *Gg*, *G. gorilla*; *Ppa*, *P. paniscus*; *Pt*, *P. troglodytes*; *Hs*, *H. sapiens*; Neg, no DNA template in the reaction. A total of eight chimpanzees, three bonobos and seven gorillas were tested with identical results. Only a subset is shown here.




Two proviruses, HERV-K105 and HERV-K110/HERV-K18 were detected in both humans and apes (Figures 2 and 3). HERV-K110 was present in humans, chimpanzees, bonobos and gorillas but not in the orang-utan (Table 1). Thus, this provirus formed after orang-utans diverged from the lineage leading to gorillas, chimpanzees, bonobos and humans, but before the latter species separated from each other. HERV-K105 was detected in humans, chimpanzees and bonobos, but not in gorillas or the orang-utan. The preintegration site, however, could not be detected in gorillas or orang-utans using several different primers based on the human sequences that flank this provirus. It is therefore unclear from this analysis whether this provirus formed after gorillas diverged from the human–chimpanzee–bonobo lineage, or if it formed earlier but was subsequently deleted in one or more lineages leading to modern apes. It is clear that at least one full-length HERV-K provirus in the human genome today has persisted since before humans, chimpanzees, bonobos and gorillas separated

**Table 1**

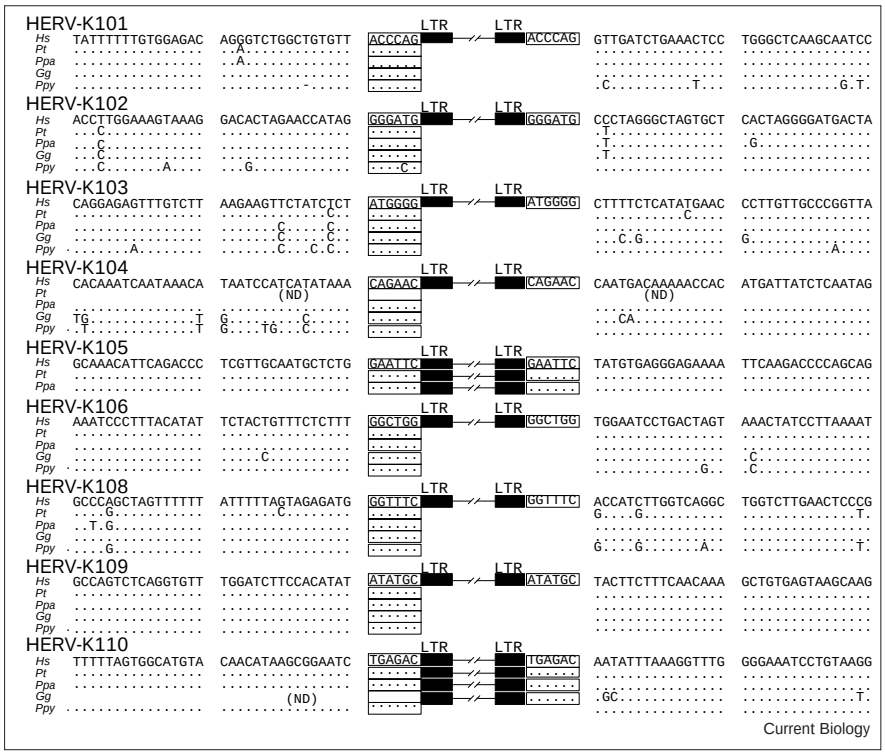
**Detection of specific HERV-K proviruses in hominids.**

| | PCR products | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Provirus | *H. sapiens* | | | *P. troglodytes* | | | *P. paniscus* | | | *G. gorilla* | | | *P. pygmaeus* | | |
| | AB* | C*D | AD | AB* | C*D | AD | AB* | C*D | AD | AB* | C*D | AD | AB* | C*D | AD |
| HERV-K101 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K102 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K103 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K104 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K105 | + | + | – | + | + | – | + | + | – | – | – | – | – | – | – |
| HERV-K106 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K107/ HERV-K10 | + | + | ND | – | – | ND | – | – | ND | – | – | ND | – | – | ND |
| HERV-K108/ HERV-K(HLM-2.HOM) | + | + | – | – | – | + | – | – | + | – | – | + | – | – | + |
| HERV-K109 | + | + | – | – | – | + | – | – | + | – | – | + | – | – | – |
| HERV-K110/ HERV-K18 | + | + | – | + | + | – | + | + | – | ND | + | – | ND | – | + |

A plus symbol indicates that a PCR product diagnostic for the particular provirus was successfully amplified using the primer combination indicated at the top of the column; a minus symbol indicates that a PCR product was not detected. ND, not determined.

HERV-K107 is integrated in a repeat unit that was detected in all human and ape samples. The PCR primers tested to date cannot distinguish the preintegration site from other loci containing repeated elements of the same family.

**Figure 3**




Nucleotide sequences of the proviral integration sites in humans and apes. Sequences are shown that flank 46 bp on either side of each provirus. The positions of the proviruses are indicated as labeled black boxes for the LTRs and thin lines for the rest of the viral genome. Cellular nucleotides that were duplicated during integration of the viral genome are outlined by rectangles. The flanking sequences from the apes are shown below the human sequences. Only nucleotides that differ in the apes are noted. Absence of a single nucleotide is indicated by a dash. Proviruses that are present in apes are indicated. ND, not determined.

during evolution, while at least eight formed after humans diverged from the extant apes.

Comparison of the sequences of the two LTRs at the ends of the genome of endogenous retroviruses allows estimation of their relative ages [1,20]. Because of the mechanism of reverse transcription, the two LTRs are very likely to be identical at the time the DNA genome integrates to form a provirus. Differences between the two LTRs of individual proviruses are likely to have occurred after establishment of each provirus in the host genome. The older the provirus, the more differences are likely to be present. On the basis of the numbers of differences between the 5′ and 3′ LTRs, HERV-K101 through HERV-K110 are listed in order of predicted age, newest first, in Table 2. The two proviruses shared by humans and the extant apes had the most differences, consistent with their being older. The LTRs of the human-specific proviruses had 0–17 differences. All but one had six or less, consistent with the idea that some of these proviruses formed relatively recently.

HERV-K genomes can be classified as type 1 or type 2 depending on whether 292 bp at the *pol–env* boundary were present or deleted, as was first observed in HERV-K10 [21]. This deletion caused a frameshift that fused Env in frame to the carboxyl portion of Pol. Of the ten proviruses analyzed here, six were type 1 and four were type 2 (Table 2). The two older proviruses present in humans and apes differ in that HERV-K105 is type 1 whereas HERV-K110 is type 2. This indicates that the 292 bp deletion occurred before the human–chimpanzee divergence. Both types of HERV-K genomes also appear to have integrated into the human genome on multiple occasions since the divergence of humans and chimpanzees.

The endogenous retroviruses that are in the human germline today are generally thought to lack infectivity because of the accumulation of mutations over evolutionary time [1]. HERV-Ks are the most likely human endogenous retroviruses to be an exception because many of the HERV-K proviruses formed in recent evolutionary times. At the time when each HERV-K provirus integrated into the human germline, all the proteins and *cis*-acting elements necessary for viral replication must have been present. To assess the coding potential of modern HERV-K proviruses, we determined the complete nucleotide sequences of seven human-specific proviruses (Figure 4). All replication-competent retroviruses except spumaviruses encode Gag–Pro–Pol and Env primary translation products. HERV-K also encodes a protein called cORF that is translated from a doubly spliced RNA and has been postulated to function in the same way as the Rev protein of the human immunodeficiency virus (HIV) [22]. Full-length open reading frames (ORFs) for the Gag–Pro–Pol and Env precursors were each present in

**Table 2**

**Properties of individual HERV-K proviruses.**

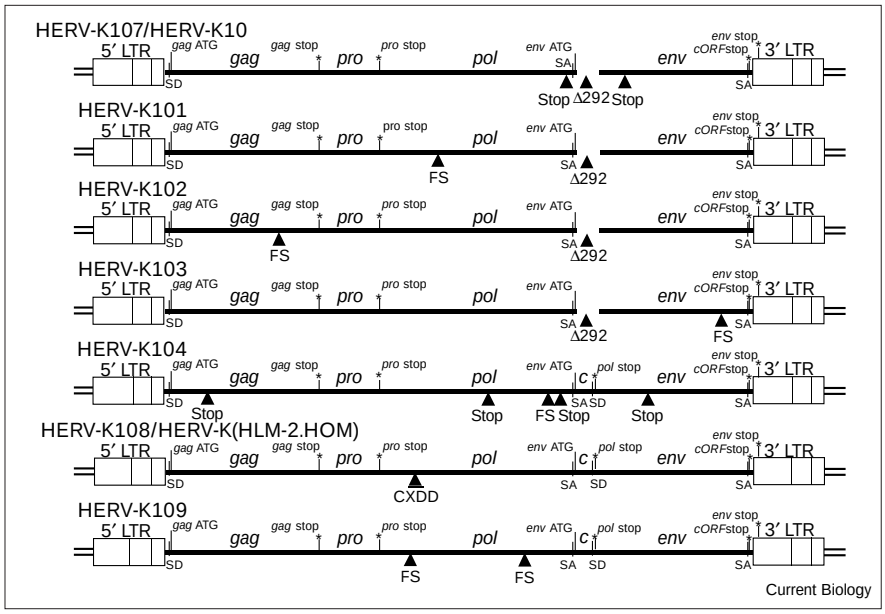
| Provirus | Number of differences between 5′ and 3′ LTRs | *env* 292 bp* | Presence of individual proviruses | | | | |
|---|---|---|---|---|---|---|---|
| | | | *H. sapiens* | *P. troglodytes* | *P. paniscus* | *G. gorilla* | *P. pygmaeus* |
| HERV-K106 | 0 | Δ | + | – | – | – | – |
| HERV-K107/ HERV-K10 | 2 | Δ | + | – | – | – | – |
| HERV-K101 | 2 | Δ | + | – | – | – | – |
| HERV-K109 | 3 | + | + | – | – | – | – |
| HERV-K102 | 4 | Δ | + | – | – | – | – |
| HERV-K103 | 6 | Δ | + | – | – | – | – |
| HERV-K108/ HERV-K(HLM-2.HOM) | 6 | + | + | – | – | – | – |
| HERV-K104 | 17 | + | + | – | – | – | – |
| HERV-K110/ HERV-K18 | 36 | + | + | + | + | + | – |
| HERV-K105 | 38 | Δ | + | + | + | ND | ND |

*The proviruses were analyzed for presence (+) or deletion (Δ) of a 292 bp sequence at the *pol–env* boundary. ND, not determined.

multiple HERV-K proviruses (Figure 4). HERV-K108/HERV-K(HLM-2.HOM) had full-length ORFs for both the Gag–Pro–Pol and Env primary translation products. A full-length ORF for Env was also present in HERV-K109. ORFs for Gag–Pro–Pol were also detected in HERV-K103 and HERV-K107/HERV-K10.

The original sequence for HERV-K10 included a frameshift in the middle of the *gag* ORF [17]. That mutation, a 1 bp deletion in HERV-K10 at position 1,750 where HERV-K107 has CTG*G*CC (base 1,750 italicized), was not present in the HERV-K107 clone. It is unclear whether the mutation was the result of a sequencing difficulty or is a genuine polymorphism among humans. Both HERV-K103 and HERV-K107/HERV-K10 are type 1 proviruses with the 292 bp deletion that spans the *pol–env* boundary. The deletion removes the carboxy-terminal 82 amino acids of Pol as defined by the common stop codon for the *pol* gene (Figure 4). That leaves 128 amino acids of Pol downstream of the conserved $DDX_{35}E$ motif

**Figure 4**




Structures of the genomes of seven human-specific HERV-K proviruses. The *gag*, *pro*, *pol*, *cORF* (c) and *env* genes are illustrated, together with the 5′ and 3′ LTRs and positions of the *gag* and *env* translational initiation codons (ATG). Asterisks, conserved stop codons; SD, splice donor site; SA, splice acceptor site; triangles, positions of premature stop codons or frameshift (FS; insertion or deletion) mutations, including the 292 bp deletion (Δ292) common to type 1 proviruses. HERV-K is predicted to utilize the –1 frameshifts [35] at the *gag* and *pro* stop codons.

(in the single-letter amino acid code; X represents any amino acid) in the integrase coding region fused to Env sequences, which is still a similar length stretch of amino acids to those in many other retroviruses [23]. The HERV-K107 clone, however, had a single base substitution at position 6,377 that created a stop codon at amino-acid residue 85 downstream of the $DDX_{35}E$ motif. That mutation was not present in HERV-K10, as translation of HERV-K10 integrase clearly demonstrated that it is fused to downstream Env amino acids [24]. Thus, the sequence at position 6,377 of the HERV-K10/HERV-K107 locus is polymorphic among modern humans. The integrase protein of HERV-K10 has been shown to have the terminal-cleavage and strand-transfer activities of retroviral integrases [24]. Two forms of HERV-K10 integrase have been shown to be active, one with a carboxyl terminus at the stop codon in *env* (Figure 4), the other with a carboxyl terminus at the boundary of *pol* and the 292 bp deletion (Figure 4) [24]. Curiously, both HERV-K103 and HERV-K10 have mutations that truncate the predicted translation product part way through Env (Figure 4 and [17]). The Gag–Pro–Pol translation product of HERV-K108/HERV-K(HLM-2.HOM) contained a substitution in the conserved YXDD sequence of the viral reverse transcriptase protein to CXDD. Those of HERV-K103 and HERV-K107 had the typical YXDD motif. All three type 2 genomes contained a full-length ORF for the cORF protein (Figure 4).

## Discussion

HERV-K genomes similar to those described here first entered the genomes of primates sometime after the platyrrhine–catarrhine divergence that occurred about 35 million years ago [25]. New HERV-K proviruses have continued to form in recent times in multiple hominoid lineages. Humans and chimpanzees diverged about five million years ago [26,27]. A HERV-K solo LTR that formed after the divergence of humans and chimpanzees has been identified in the chimpanzee genome [28]. Our data show that many of the full-length HERV-K proviruses that are in the human genome today formed subsequent to the divergence of humans and chimpanzees, although at least one predated the separation of gorillas, chimpanzees, bonobos and humans. A survey of HERV-K solo LTRs among human sequences in GenBank showed that, whereas most of these elements predated the human–chimpanzee divergence, at least nine of them entered the human genome after the divergence [29]. Depending on what fraction of the human genome was surveyed [29], there may be hundreds of HERV-K solo LTRs that are human-specific. Thus it appears that HERV-K genomes have frequently entered the germlines of hominoids in the lineage leading to modern humans during the last five million years. Most that persist today are solo LTRs but a subset has persisted as full-length proviruses.

All the ORFs and *cis*-acting sequences necessary for viral replication must have been present in the germlines of humans and other catarrhines at the times when new HERV-K genomes integrated into the human germline. The *cis*-acting sequences for viral genome transcription, nuclear export, packaging, reverse transcription and integration must have been intact in the HERV-K genomes that were the progenitors of the proviruses at new sites in the host genome. Nevertheless, the proteins necessary for viral replication could have been provided *in trans* from other loci. The Gag–Pro–Pol, cORF and Env proteins could each have been derived from different loci. It is also conceivable that HERV-K could form pseudotypes with envelopes of other viruses. The human genome encodes at least some functional HERV-K proteins [21,24,30–32], and HERV-K particles can bud from human cells [14,32–34] although there is no evidence yet that these are infectious.

Six of the seven proviruses sequenced here had at least one premature stop codon or frameshift mutation that disrupted one or more viral ORFs. Except for the 292 bp deletion, the mutations were unique to each provirus. It seems unlikely, although not formally excluded, that any single HERV-K genome has all the features necessary for replication as a viral particle. Nevertheless, it is also conceivable that complementation and/or recombination among the multiple full-length HERV-K proviruses in the human genome could lead to viral replication. ORFs for the primary translation products of the virus were present in multiple proviruses, although the Gag–Pro–Pol precursors encoded by HERV-K103, HERV-K107/HERV-K10 and HERV-K108/HERV-K(HLM-2.HOM) each had an unusual feature compared with the Gag–Pro–Pol proteins of other retroviruses. It is conceivable that the HERV-K108/HERV-K(HLM-2.HOM) provirus has all the sequences required for HERV-K replication [18], or that complementation by Gag–Pro–Pol, Env and cORF products encoded by different HERV-K proviruses might allow viral replication. A third possibility is that additional, unidentified HERV-Ks in the human genome participate in viral replication. As there are multiple full-length proviruses, recombination among HERV-K loci to assemble genomes with new combinations of viral sequences is also a possibility. Recombination or gene conversion might occur between two proviruses in the host genome, or recombination might occur during reverse transcription of two co-packaged RNA genomes if even a low level of viral replication can occur. As perhaps hundreds of HERV-K proviruses entered the human genome within the last five million years, and a subset persisted as full-length proviruses, it is possible that HERV-K is still capable of forming new proviruses in humans today.

## Conclusions

HERV-K provirus formation occurred at multiple times in the genomes of primates in the lineage leading to modern

humans. Many HERV-K genomes entered the human germline after the divergence of humans and chimpanzees. At least eight of these proviruses persisted as full-length, HERV-K genomes until the present. All the viral ORFs and *cis*-acting sequences necessary for HERV-K replication must have been intact during the recent time when these proviruses formed. Multiple, full-length ORFs for HERV-K proteins are present in the human genome today.

## Materials and methods

### *BAC and PAC screening*

Filters and clones from a human male BAC library were obtained from BACPAC Resources (Department of Human Genetics, Roswell Park Cancer Institute). Hybridization screening was performed on about 36,000 clones, giving almost twofold coverage of the human genome. Hybridizations were performed using conditions described by BACPAC Resources, except that the blocking step with human DNA was excluded, and a final wash with 1 × SSC and 0.2% SDS was performed at 65°C for 30 min. The hybridization probe was generated by random prime labeling of a fragment of the HERV-K *pol* gene that was amplified by PCR from human genomic DNA.

### *Sequencing of cellular DNA flanking cloned proviruses*

BAC DNAs were prepared as described by BACPAC Resources and sequenced directly using primers in the viral LTRs (Figure 1, primers B and C). As each provirus had two LTRs, it was necessary to eliminate the products from one of the LTRs. This was accomplished by digesting the cellular DNAs with a restriction enzyme that cleaved proviral sequences immediately adjacent to one of the LTRs, thereby prematurely terminating one of the products. Enzymes for this purpose were chosen using the sequence of the HERV-K10 provirus [17]. Cleavage adjacent to the 5′ LTR was accomplished with *Pst*I, *Ase*I, *Hin*cII, *Xcm*I or *Ear*I. *Sal*I, *Hin*cII, *Acc*I, *Xho*I or *Ban*I were used for cleavage near the 3′ LTR. Enzymes were tested individually until a readable sequence was obtained. Alternatively, BACs were digested with *Eco*RI and subcloned into a plasmid. Individual transformed colonies were screened by PCR for the presence of a 5′ or 3′ LTR. Sequencing was performed using an ABI377 apparatus at the AECOM DNA Sequencing Facility.

### *Sequencing of HERV-K proviruses*

A set of 43 PCR primers spaced at about equal intervals was designed for each strand of HERV-K using the sequence of HERV-K10 [17]. PCRs were performed on BAC templates to generate products of about 1.2–1.7 kb. Individual PCR products overlapped with those on either side. To sequence the ends of the LTRs, primers in the cellular sequences that flank each provirus (Figure 1, primers A or D) were used along with a primer in the viral genome to generate the PCR products. Each genome was completely sequenced on both strands using an ABI377 apparatus at the AECOM DNA Sequencing Facility.

### *PCR and repeated sequences*

Primers within the HERV-K genome were designed according to existing HERV-K sequences in GenBank. Primers for sequences that flank individual proviruses were based on the sequences that were determined from the BAC clones as part of these studies. Many of the HERV-K proviruses are integrated within repeated elements in the cellular genome. Attempts were made to design flanking primers that were outside of known repeated elements in the human genome. Nonetheless, the initial AD primer pairs (Figure 1) that were designed sometimes amplified members of a family of repeats from non-allelic loci when they were tested on human genomic DNAs. When that occurred, the following strategy was used to design primers that were specific for the viral loci. The sequence obtained by directly sequencing the AD PCR product amplified from human genomic DNA was compared with the sequences flanking the two sides of the cognate provirus. Nucleotide differences were identified. New PCR primers were then designed to distinguish the proviral loci from other loci that contained the repeated element by including nucleotides from the proviral loci at the 3′ ends of the primers. Such primers were then empirically tested until a pair was found for each provirus (except HERV-K107) that did amplify the product from ape DNAs but did not amplify any sequences from human genomic DNAs.

### *Accession numbers*

HERV-K sequences were deposited in GenBank with accession numbers AF095795 through AF095812, AF164609 through AF164621 and AF165228 through AF165265.

### *Supplementary material*

Supplementary material including primer sequences is available at http://current-biology.com/supmat/supmatin.htm.

## Acknowledgements

We thank A. Trubetskoy, R. Kim, B. Morrow, L. Edelmann, P. Etkind and R. Kucherlapati for advice and assistance, and S. Masseling and M. Su for participation in initial experiments. Ape tissue samples were kindly provided by Stanford University, the Yerkes Regional Primate Center, the Henry Doorly Zoo and the Milwaukee County Zoo. This work was supported by grant CA44822 from the NIH.

## References

1. Boeke JD, Stoye JP: Retrotransposons, endogenous retroviruses, and the evolution of retroelements. In *Retroviruses*. Edited by JM Coffin, SH Hughes and HE Varmus. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997:343-435.
2. Patience C, Wilkinson DA, Weiss RA: Our retroviral heritage. *Trends Genet* 1997, 13:116-120.
3. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH: Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev* 1992, 6:1457-1465.
4. Cachon-Gonzalez MB, Fenner S, Coffin JM, Moran C, Best S, Stoye JP: Structure and expression of the ***hairless*** gene of mice. *Proc Natl Acad Sci USA* 1994, 91:7717-7721.
5. Lilly F: Fv-2: identification and location of a second gene governing the spleen focus response to Friend leukemia virus in mice. *J Natl Cancer Inst* 1970, 45:163-169.
6. Best S, Le Tissier P, Towers G, Stoye JP: Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* 1996, 382:826-829.
7. Ikeda H, Sugimura H: Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. *J Virol* 1989, 63:5405-5412.
8. Golovkina TV, Chervonsky A, Dudley JP, Ross SR: Transgenic mouse mammary tumor virus superantigen expression prevents viral infection. *Cell* 1992, 69:637-645.
9. Callahan R, Drohan W, Tronick S, Schlom J: Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. *Proc Natl Acad Sci USA* 1982, 79:5503-5507.
10. Ono M: Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J Virol* 1986, 58:937-944.
11. Mariani-Costantini R, Horn TM, Callahan R: Ancestry of a human endogenous retrovirus family. *J Virol* 1989, 63:4982-4985.
12. Steinhuber S, Brack M, Hunsmann G, Schwelberger H, Dierich M, Vogetseder W: Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum Genet* 1995, 96:188-192.
13. Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu L-C, Yu CY: The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenet* 1994, 40:425-436.
14. Simpson G, Patience C, Lower R, Tonjes R, Moore H, Weiss R, Boyd M: Endogenous D-type (HERV-K) related sequences are packaged into retroviral particles in the placenta and possess open reading frames for reverse transcriptase. *Virology* 1996, 222:451-456.
15. Brown PO: Integration. In *Retroviruses*. Edited by JM Coffin, SH Hughes and HE Varmus. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997:161-203.

16. Withers-Ward ES, Kitamura Y, Barnes JP, Coffin JM: Distribution of targets for avian retrovirus DNA integration ***in vivo***. *Genes Dev* 1994, 8:1473-1487.
17. Ono M, Yasunaga T, Miyata T, Ushikubo H: Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J Virol* 1986, 60:589-598.
18. Mayer J, Sauter M, Racz A, Scherer D, Mueller-Lantzsch N, Meese E: An almost-intact human endogenous retrovirus K on human chromosome 7. *Nat Genet* 1999, 21:257-258.
19. Takahata N: Allelic genealogy and human evolution. *Mol Biol Evol* 1993, 10:2-22.
20. Dangel AW, Baker BJ, Mendoza AR, Yu CY: Complement component C4 gene intron 9 as a phylogenetic marker for primates: long terminal repeats of the endogenous retrovirus ERV-K(C4) are a molecular clock of evolution. *Immunogenet* 1995, 42:41-52.
21. Löwer R, Boller K, Hasenmaier B, Korbmacher C, Müller-Lantzsch N, Löwer J, Kurth R: Identification of human endogenous retroviruses with complex mRNA expression and particle formation. *Proc Natl Acad Sci USA* 1993, 90:4480-4484.
22. Löwer R, Tonjes R, Korbmacher C, Kurth R, Löwer J: Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. *J Virol* 1995, 69:141-149.
23. Roth MJ: Mutational analysis of the carboxyl terminus of the Moloney murine leukemia virus integration protein. *J Virol* 1991, 65:2141-2145.
24. Kitamura Y, Ayukawa T, Ishikawa T, Kanda T, Yoshiike K: Human endogenous retrovirus K10 encodes a functional integrase. *J Virol* 1996, 70:3302-3306.
25. Fleagle JG: *Primate Adaptation and Evolution,* 2nd edn. San Diego: Academic Press; 1999.
26. White T, Suwa G, Asfaw B: ***Australopithecus ramidus***, a new species of early hominid from Aramis, Ethiopia. *Nature* 1994, 371:306-312.
27. Horai S, Satta Y, Hayasaka K, Kondo R, Inoue T, Ishida T, *et al*.: Man's place in Hominoidea revealed by mitochondrial DNA genealogy. *J Mol Evol* 1992, 35:32-43.
28. Craig LC, Pirtle IL, Gracy RW, Pirtle RM: Characterization of the transcription unit and two processed pseudogenes of chimpanzee triosephosphate isomerase (TPI). *Gene* 1991, 99:217-227.
29. Medstrand P, Mager DL: Human-specific integrations of the HERV-K endogenous retrovirus family. *J Virol* 1998, 72:9782-9787.
30. Müller-Lantzsch N, Sauter M, Weiskircher A, Kramer K, Best B, Buck M, Grasser F: Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease. *AIDS Res Hum Retroviruses* 1993, 9:343-350.
31. Löwer R, Löwer J, Kurth R: The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci USA* 1996, 93:5177-5184.
32. Patience C, Simpson GR, Colletta AA, Welch HM, Weiss RA, Boyd MT: Human endogenous retrovirus expression and reverse transcriptase activity in the T47D mammary carcinoma cell line. *J Virol* 1996, 70:2654-2657.
33. Seifarth W, Baust C, Murr A, Skladny H, Krieg-Schneider F, Blusch J, *et al.*: Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J Virol* 1998, 72:8384-8391.
34. Boller K, Konig H, Sauter M, Mueller-Lantzsch N, Lower R, Lower J, Kurth R: Evidence that HERV-K is the endogenous retrovirus sequence that codes for the human teratocarcinoma-derived retrovirus HTDV. *Virology* 1993, 196:349-353.
35. Swanstrom R, Wills JW: Synthesis, assembly and processing of viral proteins. In *Retroviruses.* Edited by JM Coffin, SH Hughes and HE Varmus. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997:263-334.

---

**Because *Current Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cub – for further information, see the explanation on the contents page.**

# Supplementary material

# Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans

Madalina Barbulescu, Geoffrey Turner, Michael I. Seaman, Amos S. Deinard, Kenneth K. Kidd and Jack Lenz

Current Biology 3 August 1999, 9:861–868

**Table S1**

**DNA primers used for PCR and sequencing.**

| | | | | |
|---|---|---|---|---|
| | Flanking primers* | | | |
| Provirus | 5′-flanking primers | | 3′-flanking primers | |
| HERV-K101 | MB-1′′ | AGTCAGCAAAGGGTGGTGG | SM-4 | TCCTGAGCTGGAGAACAGACCCCA |
| HERV-K102 | MB-13 | CCCAGGACCTCTTTCTTTGTGA | MB-9 | CTTCTTCAGCATTTGAGGGTC |
| HERV-K103 | MB-2 | GATTTCGAGCCACCTCTGAAG | MB-3 | CTCAGAAACAGGCTTAAGACG |
| HERV-K104 | MB-20 | CAGCATCATCCGGACACCAAAG | MB-8 | TAGTGGTAGAGAGGGCATACT |
| HERV-K105 | MB-6 | CCTCGTGAGAAACGAACATGC | MB-32* | GATTCCAGAACACTGCTGCG |
| HERV-K106 | MB-34 | GGTGTTGCTGTGAAGGTATTC | MB-35 | TCCATGGCTATCCACGAGA |
| HERV-K107 | MB-22 | CATTACAGGTTGAGCACCCC | K10 3′fl | CCTAGTTCAAGTGATTCTCC |
| HERV-K108 | MB-27 | GTAGCTAGGGTGACAGGAGTG | MB-36 | CTTTGTGGATTGTAATTTGGGG |
| HERV-K109 | MB-23 | GTCCTTTAATGTCTCCCCTC | MB-37 | CAGATGAGATGTCAAGCAAGGT |
| HERV-K110 | MB-25 | CTCCTGAGCTTCTTGGGTCA | MB-26 | CCAGTAATGGCAATGCTGGCTATG |
| | LTR primers* | | | |
| | + Strand | | – Strand | |
| | M1 | TGTGGGGAAAAGCAAGAGAG | M4 | GTGGGTGTTTCTCGTAAGGT |
| Provirus consensus | M2 | CTGTGCTGAGGAGGATTAGT | M5 | GGACAGGCAGGAGACAGATG |
| | M3 | TCCATATGCTGAACGCTGGT | M6 | CTGAGTTGACACAGCACACG |
| | M7F | AAGCCAGGTATTGTCCAAGG | M7R | CCTTGGACAATACCTGGCTT |
| | M8F | TAAGGGAACTCAGAGGCTGG | M8R | CCAGCCTCTGAGTTCCCTTA |
| | Proviral internal primers* | | | |
| | + Strand | | – Strand | |
| | GT1F | TCTCTAGGGTGAAGGTACGC | GT16R | TTGTTCTGGAAACCATGGGC |
| | GT2F | ATGTAGCAGAGCCGGTAATG | GT17R | ATACTGAAGTTCAGCCAGCG |
| | GT3F | CCACAGTTGAGGCCAGATAC | GT18R | TTAACTGGAGGATTGGCAGC |
| | GT4F | AAGCCGGTAAGGTCATAGTG | GT19R | TAAGTCAGGTGGCTCTCTAC |
| | GT5F | CAGCCATTTGTTCCTCAGGG | GT20R | GGCACTCCAAGGAATTGAAG |
| | GT6F | GAAGGGTTGGTAGACACTGG | GT21R | GAGCGGGCATGGTGATTTCC |
| | GT7F | GTAAATCAGTGGCCGCTACC | GT22R | TTTCACAATCCTGCTCTGCC |
| Provirus consensus | GT8F | AGAGGTTGCCAATGCTGGAC | GT23R | GTAGGAATGCCTAGAGTTGG |
| | GT9F | GGTCATTCCTTCCTCACAGT | GT24R | AAGACCAATCTGCCATGCAC |
| | GT10F | ACAAGGGATGTTGAGACAGC | GT25R | GCCTGTTTCCATGTGACATC |
| | GT11F | GCAACTTGCCAAACAGGAGA | GT26R | AAGTTGCATCTGAGGAGTGG |
| | GT12F | GAGATGGTAACACCAGTCAC | GT27R | GGACTGACAGTAGGTACTTC |
| | GT12F* | AACCCATCGGAGATGCAAAG | GT28R | GGTGTGAGGCCACAGTAAGC |
| | GT13F | TGTCCAAGTGCACAAGTGAG | GT29R | GTCTCCCATCCAAATGACAG |
| | GT14F | TGGGAGGCCTCACCATCCGT | GT30R | TCTCTCTTGCTTTTCCCCAC |
| | GT15F | AGGAGTTGCTGATGGCCTCG | Δ292R | CAGCTTCTTTAGTTGTGCC |
| | GT27F | GAAGTACCTACTGTCAGTCC | | |

*Flanking primers correspond to primers A (5′) and D (3′) of Figures 1,2 and Table 1. The LTR and proviral internal primers matched the consensus sequence of the seven proviruses in Figure 4 and were spaced throughout the viral genome as needed for PCR and sequencing studies.