Effects of Recombination Rate on Human Endogenous Retrovirus Fixation and Persistence[▽]

Aris Katzourakis,* Vini Pereira,† and Michael Tristem

Department of Biology, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, United Kingdom

Received 26 February 2007/Accepted 10 July 2007

Endogenous retroviruses (ERVs) result from germ line infections by exogenous retroviruses. They can proliferate within the genome of their host species until they are either inactivated by mutation or removed by recombinational deletion. ERVs belong to a diverse group of mobile genetic elements collectively termed transposable elements (TEs). Numerous studies have attempted to elucidate the factors determining the genomic distribution and persistence of TEs. Here we show that, within humans, gene density and not recombination rate correlates with fixation of endogenous retroviruses, whereas the local recombination rate determines their persistence in a full-length state. Recombination does not appear to influence fixation either via the ectopic exchange model or by indirect models based on the efficacy of selection. We propose a model linking rates of meiotic recombination to the probability of recombinational deletion to explain the effect of recombination rate on persistence. Chromosomes 19 and Y are exceptions, possessing more elements than other regions, and we suggest this is due to low gene density and elevated rates of human ERV integration in males for chromosome Y and segmental duplication for chromosome 19.

The human genome, in common with other vertebrates, contains thousands of endogenous retroviruses (ERVs) (6, 26). The life cycle of an ERV starts with germ line infection and integration in the genome (18, 25). Only a small proportion of ERVs become fixed in the population, but they then persist until they are either overwritten by accumulated mutations, such as errors resulting from host DNA replication, or undergo recombinational deletion between their long terminal repeats (LTRs), resulting in a solo-LTR (6, 25). Within humans recombinational deletion is probably the major mechanism limiting the persistence of full-length elements (46), such that most human ERVs (HERVs) are present as solo-LTRs. HERVs are typically tens of millions of years old, and this means that the overwhelming majority of insertions are fixed within the human population (29), with only a small number of loci from the HERV-K(HML2) family being insertionally polymorphic in different human individuals (3).

The distribution of HERVs within the genome will reflect the likelihood of integrations becoming fixed in different genomic locations. Several studies, including those investigating transposable elements (TEs) in *Drosophila* and long interspersed nuclear elements in humans have implicated recombination in determining TE fixation (2, 8). For example, lower levels of TE induced ectopic exchange (where homologous TEs in different genomic locations recombine, which can lead to major genomic rearrangements) are expected in regions of reduced recombination, thus leading to the accumulation of TEs (30, 52). Local recombination rates could also influence

fixation indirectly, based on models that assume that insertions are slightly deleterious (14, 36), because the efficacy of selection in the host genome increases with recombination rate. Gene density and GC content have also been implicated previously in exerting selection pressure against the fixation of HERVs and other TEs (34, 53). TE insertions, either directly into genes or in close proximity to them, would usually be highly deleterious since they would disrupt normal gene expression (53). Occasionally, such insertions may be beneficial, for example, by providing promoters and enhancers that are co-opted by the host for the regulation of gene expression (13, 50, 51).

In the present study, we use retroviral sequence data, including solo-LTRs, from the human genome sequence to understand factors influencing two major stages of the HERV life cycle: (i) from initial integration to fixation and (ii) the persistence of fixed, full-length HERVs over time. An overabundance of HERVs, detected within chromosome Y and a 5-Mb section of chromosome 19, is also discussed.

MATERIALS AND METHODS

Molecular paleontology of HERVs. We ran RepeatMasker (45) on the human genome sequence (build 31) in order to retrieve the coordinates of genomic regions homologous to consensus sequences of 31 HERV families, as previously described (26). In order to identify “solo-LTRs” and “full-length” HERVs, we processed the RepeatMasker annotation of the human genome with the program LTR_MINER (37). Here we use the term “full-length HERV” in a structural sense to denote a provirus that retains at least part of both its LTRs and its internal region (IR). Processing by LTR_MINER was required because RepeatMasker annotates the LTRs and IR of a given proviral insertion as separate repeats, which need to be linked for the identification of a full-length HERV. Moreover, after its insertion the LTR or IR sequence of a given provirus may become fragmented by indels; in this situation the different fragments of an LTR or an IR may be annotated by RepeatMasker as separate repeats, especially if interrupted by subsequent insertions.

LTR_MINER is able to defragment LTRs and IRs separately and then link sets of two defragmented LTRs and one defragmented IR into the same full-length HERV (i.e., the inference that the defragmented sequences originated in the same proviral insertion). This defragmentation procedure is also crucial for

* Corresponding author. Mailing address: Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom. Phone: (44) 1865-281258. Fax: (44) 1865-271249. E-mail: aris.katzourakis@zoo.ox.ac.uk.

† Present address: Centre for the Study of Evolution, School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, United Kingdom.

[▽] Published ahead of print on 18 July 2007.

the identification of solo-LTRs, since we use this term exclusively to denote the result of a recombinational deletion event (46), as opposed to LTRs that have become separated from the IR due to an intervening insertion of an unrelated sequence.

Defragmentation of multiple similarity hits. LTR_MINER parameters were set so that sequence similarity hits to the same element family were defragmented only when they had the same orientation on the chromosome, and their genomic order and sequences mapped onto consecutive (though not necessarily contiguous) regions of the respective family consensus sequence. Occasionally, inferred consecutive fragments overlapped slightly; LTR_MINER pieces together such fragments, provided that their overlap does not exceed 40 bp.

Identification of full-length elements. After the defragmentation of all hits to LTRs, and separately of hits to IRs, in the original RepeatMasker annotation, LTR_MINER identifies full-length HERVs by searching for their structural pattern: LTR-IR-LTR (a defragmented LTR followed by a defragmented IR followed by a defragmented LTR of the same family and orientation). Note that here the defragmented LTR or IR may consist of multiple RepeatMasker hits; furthermore, the LTR-IR boundaries need not be contiguous (i.e., the boundaries may be separated by an unrelated insertion). However, only patterns for which the distance between the inner ends of the two LTRs is less than 30 kbp are considered candidate full-length HERVs. Finally, in order to check whether the pattern could be straddling a nested insertion of the same family, the search is then recursively extended from each end of the pattern for further hits to an IR and an LTR (of same family and orientation), subject to the constraint on the maximum distance between intra-element LTRs mentioned above. The two LTRs of the innermost pattern are classified as a pair of intra-element LTRs (and in this way any nested insertions of the same family could be identified). The automatic annotations produced by LTR_MINER have been verified by comparison to manually curated datasets with multiple levels of nested insertions (V. Pereira, unpublished data).

Identification of solo-LTRs. LTR_MINER was set to classify a defragmented LTR as a "solo-LTR" if no other LTR or IR sequence (of same family and in the same orientation) is present within a 5-kbp radius from the fragment's ends. The aim was the identification of elements resulting from deletion (of the IR and one LTR) events via homologous recombination between intra-element LTRs and not to classify as solo-LTRs sequences that are separated from IRs because of insertions.

Gene density data. Data on the number of genes, total amount of sequence data, and chromosome length for each chromosome were taken from build 31 of the human genome held at the National Centre for Biotechnology Information (freely downloadable from <http://www.ncbi.nlm.nih.gov>).

Determination of local recombination rate. We used the high-resolution recombination map of Kong et al. (28) to compute the recombination rates at HERV-containing loci. Kong et al. (28) genotyped 5,136 microsatellite markers for 146 Icelandic families, with a total of 1,257 meiotic events. This map has greater accuracy and about fivefold-higher resolution than previous genetic maps (9, 12). The markers were mapped on to an updated version of the August 2001 freeze of the Human Genome Project Working Draft (29). The published recombination rates represent the average for a window of 3 Mb, centered on the marker.

Every intact HERV and solitary LTR identified by using LTR_MINER was assigned a recombination rate based on the recombination rate of the nearest marker. Locations of recombination markers and HERVs were taken from Kong et al. (28) and the LTR_MINER output, respectively. Only elements within 1.5 Mb of a marker were included in this analysis. The process of assigning HERVs to the nearest marker was automated by using Python scripts (available on request). Since HERV integration is random with respect to genomic location, the exclusion of some elements beyond the range of these markers would not be expected to result in any bias. A similar approach has been used to investigate the effect of local recombination rates on divergence between the human and chimpanzee (21). Recombination rates were divided into evenly sized bins of 0.1 cM/Mb for rates between 0 and 3 centimorgans (cM)/Mb, resulting in 30 bins, plus a high recombination rate bin for HERVs at loci with recombination rates above 3 cM/Mb. HERVs at loci with higher recombination rates were pooled into a single bin due to lower copy numbers. The high recombination rate bin included recombination rates ranging from >3 to 7.27 cM/Mb, with a mean of 3.56 cM/Mb. The precise x coordinate of every recombination bin was calculated as the average of the recombination values of both full-length HERVs and solo-LTRs within each recombination bin. The significance of the relationship between the paired/solitary LTR ratio and recombination rate was assessed by using linear regression. In addition, we also performed a binomial regression where every HERV was described as being either a solo-LTR or a full-length HERV. The predictor variable was the local recombination rate. Local recom-

ination rate was treated as a continuous variable in this analysis, using the recombination rate of the nearest marker rather than the recombination rate bins of the linear regression.

The relationship between the total number of HERV integration events (solitary LTRs plus complete elements) and local recombination rate was examined by adding the number of full-length HERVs and solo-LTRs at each recombination bin and dividing by the total amount of sequenced nucleotides that each bin covers. This was computed by adding the full length of the genome covered by markers corresponding to each bin and then subtracting regions of the genome for each bin that have not yet been sequenced.

Test for differences in HERV density in chromosome Y and 19p12. Differences in HERV density in the Y chromosome and in a 5-Mb region of chromosome 19 compared to the rest of the genome were assessed by using a 2-by-2 χ^2 contingency table. Chromosome 19 was split into two, because although it has the highest gene density of all of the human chromosomes (29), it contains a 5-Mb region on the p arm near the centromere that is relatively gene poor. This region is at locus 19p12. The observed and expected number of elements on chromosome Y and in the remainder of the genome were compared. The expected frequencies were calculated assuming that the total number of elements was distributed between Y and the rest of the genome relative to the total amount of DNA in each partition of the genome. This test was performed both for the number of full-length HERVs and for the total number of integrations. Comparisons were made both between chromosome Y and the rest of the genome and between the 5-MB region of locus 19p12 and the rest of the genome.

RESULTS AND DISCUSSION

We used the program LTR_MINER (37) to identify solo-LTRs and full-length HERVs (defined as having at least a portion of both LTRs and an internal sequence) within the human genome sequence and recorded their location. We identified 2,375 HERVs and 41,164 solo-LTRs. Thus, ca. 95% of HERV insertions are represented by solo-LTRs.

HERV fixation is independent of local recombination rate. We used the sum of solo-LTRs and full-length HERVs to measure fixation in different genomic regions. This is in contrast to previous reports that used overall retroviral content (without assigning retroviral DNA to particular insertion events) rather than a direct measure of the number of HERVs insertions that have become fixed (34). The inclusion of solo-LTRs removes confounding effects caused by potential differences in the rate of full-length HERV loss (via recombinational deletion) in different genomic locations. We used multiple regression analysis to explicitly compare the association of both gene density and recombination rate (which are both correlated to GC content [28]) to HERV fixation, between entire chromosomes. We found that gene density and not recombination rate is the significant predictor of HERV fixation ($P < 0.0001$ and $P = 0.2$ in the multiple regression, respectively). Thus, consistent with previous reports (34), gene density correlates with HERV fixation; Fig. 1A shows a negative correlation between the density of total fixation events and gene density among the chromosomes ($P < 0.0001$). Chromosomes Y and 19 are outliers, containing more HERVs than would be expected from their gene densities (discussed below).

We also investigated the effect of local rates of recombination on HERV fixation on a finer scale, again using the sum of solo-LTRs and full-length HERVs. This is because local recombination rates vary to a much greater extent than chromosome-averaged recombination rates (28), and the latter approach may lack power to identify local effects of recombination on fixation. These data were plotted against local (meiotic) recombination rate, based on 5,136 microsatellite markers, with each value representing the average recombination

FIG. 1. (A) Scatterplot showing the density of the total number of HERV integrations (solo-LTRs plus paired LTRs) against gene density. Filled diamonds represent human chromosomes, whereas the open diamond represents the 5-Mb region at 19p12 (note that the chromosome 19 diamond does not include 19p12). The trendline and R^2 do not include data from chromosomes Y, 19, and 19p12. (B) Scatterplot showing the total fixation density for different host recombination rates. The recombination rate was divided into equally sized bins of 0.1 cM/Mb for values between 0 and 3 cM/Mb, with a single recombination rate bin for values of >3 cM/Mb. The fixation density was calculated by dividing the total number of fixations corresponding to each recombination bin, divided by the total amount of sequenced DNA in that bin. (C) Scatterplot showing the ratio of paired to solo-LTRs against recombination rate. The recombination rate was calculated as in panel B.

rate for a 3-Mb window centered on the marker (28). Only HERVs (numbering 2,146) and solo-LTRs (numbering 38,234) that were within 1.5 Mb of the markers were included.

We find no evidence that host recombination rate correlates with fixation (Fig. 1b, $P = 0.49$), despite a 30-fold variation in recombination rate, from 0.1 to 3 cM/Mb. This is contrary to predictions suggested by both the ectopic exchange model and models that implicate recombination indirectly with the probability of fixation (14, 16). Furthermore, our data are consistent with results obtained from TEs in *Arabidopsis thaliana* (53) and retrotransposons in *Caenorhabditis elegans* (14), which show no relationship between element density and local recombination rate (Table 1).

Local recombination rate correlates with HERV persistence.

Previous reports investigating TE distribution have examined correlates of TE density and by implication the factors that influence TE fixation rather than persistence (2, 8, 14, 53). HERVs, like LTR retrotransposons, can undergo recombinational deletion resulting in a solo-LTR. The extent of recombinational deletion at any genomic location can be determined by calculating the ratio of full-length HERVs (with both LTRs) to their solitary LTR counterparts. This method therefore allows the persistence of HERVs in their full-length state to be investigated in detail.

We plotted the ratio of paired to solo-LTRs against local recombination rate (Fig. 1C). We found, in contrast to total fixation events (Fig. 1B), a highly significant relationship between the paired/solo-LTR ratio and the local recombination rate ($P < 0.0001$), with a threefold increase in recombination rate halving the ratio. The local recombination rate has therefore played a major role in limiting the persistence of HERVs in their full-length state.

It is likely that HERVs integrate at random with respect to the local rate of recombination. Although integration targeting has been demonstrated for some exogenous retroviruses, the integration preferences of such viruses does not correlate with the genomic distribution of related ERVs (1, 41). This suggests that integration preference does not determine the distribution of ERVs in the genome. The probability of ERV fixation is also independent of the rate of recombination because (as shown above) gene density, not recombination rate, influences the probability of fixation. Once fixed, the persistence of the element in its full-length state then depends on the local recombination rate. Full-length elements situated in regions of low recombination undergo less deletion and hence persist for a greater period of time. We propose that this is because the rate of intrastrand homologous recombination within a genome (the mechanism generating solo-LTRs) (46) is linked by as-yet-unknown mechanistic processes to the background rate of meiotic recombination. Thus, the likelihood of solo-LTR formation correlates with the probability of a meiotic recombination event at any given location.

How does this model differ from other previously proposed scenarios such as the ectopic exchange (30, 52), gene disruption (17, 33), or deleterious gene product (36) models? Although the ectopic exchange model is consistent with the increased solo-LTR ratio in areas of high recombination (Fig. 1C) (because ectopic exchange occurs more frequently between longer, full-length elements (8), the solo-LTRs would still be deleterious to some extent. However, this is not the

TABLE 1. Different studies of determinants of TE fixation and persistence^a

Study type and host species	Transposable element type(s)	Factor investigated	Significant effect	Study
Studies investigating factors affecting TE fixation				
<i>Drosophila melanogaster</i>	Range of retrotransposons and DNA transposons	Recombination rate	Yes	Bartolomé et al. (2)
<i>Homo sapiens</i>	LINE elements	Recombination rate	Yes	Boissinot et al. (8)
<i>Caenorhabditis elegans</i>	DNA transposons	Recombination rate	Yes ^b	Duret et al. (14)
<i>Arabidopsis thaliana</i>	Range of retrotransposons and DNA transposons	Recombination rate	No	Wright et al. (53)
<i>Caenorhabditis elegans</i>	Retrotransposons	Recombination rate	No	Duret et al. (14)
<i>Homo sapiens</i>	HERVs	Recombination rate	No	This study
<i>Homo sapiens</i>	HERVs	Gene density	Yes	Medstrand et al. (34)
<i>Arabidopsis thaliana</i>	Range of retrotransposons and DNA transposons	Gene density	Yes	Wright et al. (53)
<i>Homo sapiens</i>	HERVs	Gene density	Yes	This study
<i>Drosophila melanogaster</i>	Range of retrotransposons and DNA transposons	Gene density	No	Bartolomé et al. (2)
Studies investigating factors affecting TE persistence				
<i>Homo sapiens</i>	HERVs	Recombination rate	Yes	This study

^a The factors considered included gene density, due to the deleterious effects of TE insertion, and recombination rate, both under the ectopic exchange model and also indirect models based on the efficacy of selection.

^b The result is a positive association, contrary to the prediction of the ectopic exchange model.

case: fixation frequency does not depend on the local recombination rate, which contrasts with the expectation of the ectopic exchange model (Fig. 1B).

Under the gene disruption model (17, 33), solo-LTRs would also be expected to be deleterious, since they contain the retroviral regulatory sequences affecting host gene expression. Our data support this model in terms of fixation, since we note an underrepresentation of fixed elements in gene-rich regions. The gene disruption model is unlikely to apply to persistence (Fig. 1C) since both solo-LTRs and full-length HERVs are deleterious (retroviral promoter and enhancer sequences are in the LTR). Although the formation of a solo-LTR can, in certain cases, lead to a reversion of a somatic mutation (43, 47), there are also cases of solo-LTRs disrupting the expression or splicing of nearby genes (24, 42). Thus, it seems likely that many solo-LTRs present within or near genes would be deleterious and would therefore be rapidly lost by selection.

There are, however, scenarios under which solo-LTRs may confer a selective advantage to the host compared to full-length HERVs. Under the deleterious gene product model (36), a solo-LTR does not express retroviral genes or form new viral particles that could be deleterious. Figure 1C is theoretically consistent with this model, since full-length elements are preferentially removed in regions of high recombination. However, it is our opinion that (at least within humans) negative selection would act quickly against intact and active viruses, largely preventing them from reaching fixation within the host population (3). This is consistent with the presence of replication competent viruses in other mammals, most of which are unfixed within the host species (10, 11, 19, 35, 38). In particular, the majority of active and infectious endogenous murine retroviruses are present in only a few inbred strains of mice (6). During this period (when the LTRs are largely identical), we have shown that recombinational deletion occurs extremely rapidly (5), and this masks the effect of elevated levels of

recombinational deletion according to local recombination rate. However, as LTR divergence increases over time due to host induced mutation, the rate of recombinational deletion greatly slows and the effects of local recombination rate then become more pronounced and important in determining long-term persistence. Thus, few, if any, active HERVs contribute to the data shown in Fig. 1C, which, instead, represent the mechanistic generation of solo-LTRs via intra-strand homologous recombination rather than by any selection driven process.

HERVs are overrepresented on chromosome 19 and Y. Previous reports have noted an overrepresentation of HERVs on chromosomes 19 and Y (34), a pattern that is repeated when solo-LTRs are included (Fig. 1A and Table 2). The overrepresentation of HERVs on chromosome 19 is largely due to a 5-Mb span at locus 19p12, a region containing features of both euchromatin and heterochromatin (22). Table 2 shows that full-length elements are about eightfold overrepresented in

TABLE 2. Observed and expected numbers of full-length HERVs and total integrations

Locus	Type	No. of HERVs		χ^2	P
		Observed	Expected ^a		
Chromosome Y	Full-length HERVs	75	18.6	33.47	<0.0001
	Total integrations	739	342.35	146.63	<0.0001
19p12	Full-length HERVs	30	4.049	18.41	<0.0001
	Total integrations	190	75.033	49.17	<0.0001
Chromosome 19	Full-length HERVs	44	41.23	0.038	0.85
	Total integrations	953	773.23	18.9	<0.0001

^a The expected values were calculated based on the proportion of the total amount of sequenced DNA that each category represents and represent genome-wide average densities of HERVs (excluding the chromosomes that are tested for a significant difference). Significance was assessed by χ^2 . Locus 19p12 was subtracted from the data for chromosome 19.

this region, whereas the total number of elements (solo-LTRs plus full-length elements) are approximately threefold over-represented. Thus, there are both more integration events and a higher proportion of full-length elements than expected. 19p12 is rich in zinc-finger (ZNF) genes that appear to have arisen via segmental duplication over the past 50 million years, probably via unequal crossover between β -satellite repeats (15). There are numerous solo-LTRs in the close vicinity of these ZNF genes, which are derived from only a small number of HERV families. BLAST analysis using one copy from each solo-LTR type as a probe revealed that, in each case, almost all of the top matches were from 19p12, suggesting that many solo-LTRs in this region arose as a by-product of the ZNF gene cluster expansion (unpublished data). To investigate this further, we reconstructed the phylogeny of one of the solo-LTR types present at high copy number in 19p12. The maximum-likelihood phylogeny was estimated under the HKY model of substitution (with the parameters of the model also estimated from the data) in PAUP* (49). This showed that the chromosome 19 LTRs clustered together to the exclusion of most LTRs present on other chromosomes (not shown), which is consistent with their proliferation via segmental duplication. Although the total number of integrations on 19p12 can therefore probably be explained by segmental duplication, the proportion of full-length elements to solo-LTRs cannot. Indeed, by showing that many of the solo-LTRs in this region probably arose by segmental duplication, the proportion of full-length elements is skewed further. It remains unclear why this should be the case.

Chromosome Y has approximately twice the expected number of fixed elements (Fig. 1A). This overrepresentation has previously been attributed to a low recombination rate, high heterochromatic composition, or higher tolerance for new integrations (27, 29, 34). We rule out an effect of recombination, since recombination does not affect fixation density. Furthermore, since only euchromatic regions of chromosome Y have been sequenced to date, its heterochromatic nature is not responsible. The observed overrepresentation of elements can only partly be accounted for by the lower gene density of chromosome Y (i.e., a higher tolerance for new integrations) and is above that expected even given its gene poor nature (Fig. 1A). This suggests that another factor may also be influencing fixation.

We hypothesize that the rate of HERV fixation is higher in males, as a result of a higher underlying integration frequency in the male germ line. Most exogenous retroviruses require dividing cells in order to cross the nuclear membrane and thereby integrate into the host genome (39). The lentiviruses (e.g., human immunodeficiency virus) represent a notable exception to this general rule by infecting nondividing cells (31). Other viruses, such as avian leukosis virus, are intermediate between these two states, being able to infect nondividing cells, albeit less efficiently than human immunodeficiency virus (48). Many HERVs replicate via the same mechanism as exogenous retroviruses (4) and so would also be expected to require dividing cells in order to integrate. Since males require a large number of germ line cell divisions to produce sperm from spermatogonial stem cells (44), the male germ line may be a more suitable environment for HERV replication and integration than the female germ line, where germ line cell division

ceases during fetal development. Because the Y chromosome resides only in males, it should therefore acquire more integrations than the somatic chromosomes, which spend half of their time in females. In line with this hypothesis, we note that the male mutation rate is approximately twice that of females (7, 29), approximately equal to the excess of HERVs on chromosome Y. It is thought that the higher mutation rate in males is due to the higher number of germ line cell divisions per generation in males, a process similar to the one we propose (7, 20, 44). In contrast to this, *in vivo* experiments with mice have noted more effective transmission in the female than in the male germ line (23, 32, 40). However, these experiments were performed on highly inbred strains of mice, and it remains to be determined whether the same situation occurs with normal outbreeding over longer timescales. Comparison of ERV fixation between autosomes and sex chromosomes in a range of other species should allow us to test the validity and generality of the male-biased proliferation hypothesis.

Concluding remarks. We have shown here that, within humans, the genomic distribution and persistence of ERVs is controlled by gene density and local recombination rate, respectively. This is consistent with certain TE classes in some species (14, 34, 53) but not others (2, 14) (Table 1). Thus, factors controlling the evolution, distribution, and persistence of TEs are likely to depend on a complex interplay between specific TE classes and particular host species. The analysis of sets of TEs from a wide variety of host genome sequences should eventually allow general trends in TE evolution to be separated from factors specific to the TE or host life history.

ACKNOWLEDGMENTS

We thank R. Belshaw and A. Burt for helpful discussions. We also thank the referees for helpful comments.

The MRC, NERC, and Wellcome Trust provided financial support.

REFERENCES

1. Barr, S. D., J. Leipzig, P. Shinn, J. R. Ecker, and F. D. Bushman. 2005. Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *J. Virol.* **79**:12035–12044.
2. Bartolomé, C., X. Maside, and B. Charlesworth. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* **19**:926–937.
3. Belshaw, R., A. L. A. Dawson, J. Woolven-Allen, J. Redding, A. Burt, and M. Tristem. 2005. Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J. Virol.* **79**:12507–12514.
4. Belshaw, R., V. Pereira, A. Katzourakis, G. Talbot, J. Paces, A. Burt, and M. Tristem. 2004. Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* **101**:4894–4899.
5. Belshaw, R., J. Watson, A. Katzourakis, A. Howe, J. Woolven-Allen, A. Burt, and M. Tristem. 2007. Rate of recombinational deletion among human endogenous retroviruses. *J. Virol.* **81**:9437–9442.
6. Boeke, J. D., and J. P. Stoye. 1997. Retrotransposons, endogenous retroviruses, and the evolution of retroelements, p. 343–435. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
7. Bohossian, H. B., H. Skaletsky, and D. C. Page. 2000. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. *Nature* **406**:622–625.
8. Boissinot, S., A. Entezam, and A. V. Furano. 2001. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18**:926–935.
9. Broman, K. W., J. C. Murray, V. C. Sheffield, R. L. White, and J. L. Weber. 1998. Comprehensive human genetic maps: individual and sex-specific variation in recombination. *Am. J. Hum. Genet.* **63**:861–869.
10. Carlson, J., M. Lyon, J. Bishop, A. Vaiman, E. Cribiu, J. F. Mornex, S. Brown, D. Knudson, J. DeMartini, and C. Leroux. 2003. Chromosomal distribution of endogenous jaagsiekte sheep retrovirus proviral sequences in the sheep genome. *J. Virol.* **77**:9662–9668.

11. DeMartini, J. C., J. O. Carlson, C. Leroux, T. Spencer, and A. Palmirini. 2003. Endogenous retroviruses related to jaagsiekte sheep retrovirus. *Jaagsiekte Sheep Retrovirus Lung Cancer* **275**:117–137.
12. Dib, C., S. Faure, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette, and J. Weissenbach. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**:152–154.
13. Dunn, C. A., P. Medstrand, and D. L. Mager. 2003. An endogenous retroviral long terminal repeat is the dominant promoter for human beta 1,3-galactosyl-transferase 5 in the colon. *Proc. Natl. Acad. Sci. USA* **100**:12841–12846.
14. Duret, L., G. Marais, and C. Biemont. 2000. Transposons but not retrotransposons are located preferentially in regions of high recombination rate in *Caenorhabditis elegans*. *Genetics* **156**:1661–1669.
15. Eichler, E. E., S. M. Hoffman, and A. A. Adamson. 1998. Complex beta-satellite repeat structures and the expansion of the zinc finger gene cluster in 19p12. *Genome Res.* **8**:791–808.
16. Eickbush, T. H., and A. V. Furano. 2002. Fruit flies and humans respond differently to retrotransposons. *Curr. Opin. Genet. Dev.* **12**:669–674.
17. Finnegan, D. J. 1992. Transposable elements, p. 1096–1107. *In* D. L. Lindsley and G. Zimm (ed.), *The genome of Drosophila melanogaster*. Academic Press, Inc., New York, NY.
18. Gifford, R., and M. Tristem. 2003. The evolution, distribution, and diversity of endogenous retroviruses. *Virus Genes* **26**:291–316.
19. Gorbovitskaia, M., Z. L. Liu, N. Bourgeois, N. Li, Z. X. Lian, P. Chardon, and C. Rogel-Gaillard. 2003. Characterization of two porcine endogenous retrovirus integration loci and variability in pigs. *Immunogenetics* **55**:262–270.
20. Haldane, J. B. S. 1947. The mutation rate of the gene for haemophilia, and its segregation ratios in males and females. *Ann. Eugenics* **13**:262–271.
21. Hellmann, I., I. Ebersberger, S. E. Ptak, S. Paabo, and M. Przeworski. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* **72**:1527–1535.
22. Horvath, J. E., J. A. Bailey, D. P. Locke, and E. E. Eichler. 2001. Lessons from the human genome: transitions between euchromatin and heterochromatin. *Hum. Mol. Genet.* **10**:2215–2223.
23. Jenkins, N. A., and N. G. Copeland. 1985. High-frequency germline acquisition of ecotropic MULV proviruses in Swr/J-Rf/J hybrid mice. *Cell* **43**:811–819.
24. Jun, K., S. B. Lee, and H. S. Shin. 2000. Insertion of a retroviral solo long terminal repeat in *mdr-3* locus disrupts mRNA splicing in mice. *Mammalian Genome* **11**:843–848.
25. Katourakis, A., A. Rambaut, and O. G. Pybus. 2005. The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol.* **13**:463–468.
26. Katourakis, A., and M. Tristem. 2005. Phylogeny of human endogenous and exogenous retroviruses, p. 186–203. *In* E. D. Sverdlov (ed.), *Retroviruses and primate genome evolution*. Landes Bioscience, Georgetown, TX.
27. Kjellman, C., H. O. Sjögren, and B. Widegren. 1995. The Y chromosome: a graveyard for endogenous retroviruses. *Gene* **161**:163–170.
28. Kong, A., D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, and K. Stefansson. 2002. A high-resolution recombination map of the human genome. *Nat. Genet.* **31**:241–247.
29. Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaanty, T. L. Miner, A. Delehaanty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
30. Langley, C. H., E. Montgomery, R. Hudson, N. Kaplan, and B. Charlesworth. 1988. On the role of unequal exchange in the containment of transposable element copy number. *Genetic Res.* **52**:223–235.
31. Lewis, P. F., and M. Emerman. 1994. Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus. *J. Virol.* **68**:510–516.
32. Lock, L. F., E. Keshet, D. J. Gilbert, N. A. Jenkins, and N. G. Copeland. 1988. Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice. *EMBO J.* **7**:4169–4177.
33. McDonald, J. F., L. V. Matyunina, S. Wilson, I. K. Jordan, N. J. Bowen, and W. J. Miller. 1997. LTR retrotransposons and the evolution of eukaryotic enhancers. *Genetica* **100**:3–13.
34. Medstrand, P., L. N. van de Lagemaat, and D. L. Mager. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* **12**:1483–1495.
35. Niebert, M., and R. R. Tonjes. 2003. Molecular cloning and functional characterization of infectious PERV and development of diagnostic tests. *Xeno-Transplantation* **278**:217–237.
36. Nuzhdin, S. V. 1999. Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* **107**:129–137.
37. Pereira, V. 2004. Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol.* **5**:R79.
38. Roca, A. L., J. Peacon-Slatery, and S. J. O'Brien. 2004. Genomically intact endogenous feline leukemia viruses of recent origin. *J. Virol.* **78**:4370–4375.
39. Roe, T. Y., T. C. Reynolds, G. Yu, and P. O. Brown. 1993. Integration of murine leukemia-virus DNA depends on mitosis. *EMBO J.* **12**:2099–2108.
40. Rowe, W. P., and C. A. Kozak. 1980. Germ-line reinsertions of Akv murine leukemia-virus genomes in Akv-1 congenic mice. *Proc. Natl. Acad. Sci. USA* **77**:4871–4874.
41. Schroder, A. R. W., P. Shinn, H. M. Chen, C. Berry, J. R. Ecker, and F. Bushman. 2002. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**:521–529.
42. Seperack, P. K., J. A. Mercer, M. C. Strobel, N. G. Copeland, and N. A. Jenkins. 1995. Retroviral sequences located within an intron of the dilute gene alter dilute expression in a tissue-specific manner. *EMBO J.* **14**:2326–2332.
43. Seperack, P. K., M. C. Strobel, D. J. Corrow, N. A. Jenkins, and N. G. Copeland. 1988. Somatic and germ-line reverse mutation rates of the retrovirus-induced dilute coat-color mutation of DBA mice. *Proc. Natl. Acad. Sci. USA* **85**:189–192.
44. Shimmin, L. C., B. H. J. Chang, and W. H. Li. 1993. Male-driven evolution of DNA sequences. *Nature* **362**:745–747.
45. Smit, A. F. A., R. Hubley, and P. Green. 2004. RepeatMasker Open-3.0. [Online.] <http://www.repeatmasker.org>.
46. Stoye, J. P. 2001. Endogenous retroviruses: still active after all these years? *Curr. Biol.* **11**:R914–R916.
47. Stoye, J. P., S. Fenner, G. E. Greenoak, C. Moran, and J. M. Coffin. 1988. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* **54**:383–391.
48. Suzuki, Y., and R. Craigie. 2007. The road to chromatin: nuclear entry of retroviruses. *Nat. Rev. Microbiol.* **5**:187–196.
49. Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (* and other methods), 4th ed. Sinauer Associates, Sunderland, MA.
50. Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson, and M. H. Meisler. 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6**:1457–1465.
51. van de Lagemaat, L. N., J. R. Landry, D. L. Mager, and P. Medstrand. 2003. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* **19**:530–536.
52. Virgin, J. B., and J. P. Bailey. 1998. The M26 hotspot of Schizosaccharomyces pombe stimulates meiotic ectopic recombination and chromosomal rearrangements. *Genetics* **149**:1191–1204.
53. Wright, S. I., N. Agrawal, and T. E. Bureau. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res.* **13**:1897–1903.