Virology **279**, 558–569 (2001)
doi:10.1006/viro.2000.0737, available online at http://www.idealibrary.com on IDEAL®

metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Characterization of the Three HERV-H Proviruses with an Open Envelope Reading Frame Encompassing the Immunosuppressive Domain and Evolutionary History in Primates

Nathalie de Parseval,* Jean-François Casella,* Laetitia Gressin,† and Thierry Heidmann*[,1]

**Unité des Rétrovirus Endogènes et Éléments Rétroïdes des Eucaryotes Supérieurs, CNRS UMR 1573, Institut Gustave Roussy, 39 rue Camille Desmoulins, Villejuif Cedex 94805, France; and †Centre d'Etude du Polymorphisme Humain, 27 rue Juliette Dodu, Paris 75010, France*

Received August 10, 2000; returned to author for revision October 3, 2000; accepted October 31, 2000

The HERV-H family is one of the largest human endogenous retrovirus families, with approximately 1000 elements. Using a direct coupled *in vitro* transcription/translation approach (PTT for protein truncation test) and an extended series of primers on human genomic DNA, on monochromosomal hybrids and on a BAC library, we could demonstrate that there are only three envelopes with a large open reading frame encompassing the immunosuppressive (ISU) domain, corresponding to 62-, 60-, and 59-kDa potential translational products. The associated proviruses, HERV-H/env62, HERV-H/env60, and HERV-H/env59 were sequenced together with their flanking DNA and mapped by FISH, and their entry times within the primate lineage were determined. Analysis of the LTR sequences revealed numerous recombinational and/or homogenization events in the course of evolution, with divergences between 5′ and 3′ LTRs higher than expected for a simple time-dependent genetic drift. PTT analyses further revealed that the three large envelopes in humans are prematurely stopped in the majority of primates, and sequencing of the largest envelope gene, from HERV-H/env62, in five human individuals revealed two polymorphic sites. The results are consistent with the absence of a strong selective pressure for the conservation of a functional envelope gene of possible benefit for the host, but do not exclude somatic effects possibly associated with the immunosuppressive domain carried by these genes. © 2001 Academic Press

INTRODUCTION

Approximately 1% of the human genome consists of proviral structures called human endogenous retroviruses (HERVs). These proviruses most probably originate from primary infections of germline cells by ancient retroviruses. Thereafter, they were replicated as host genes and, in some cases, amplified by retrotransposition and/or reinfection in the course of evolution (reviewed in Löwer *et al.,* 1996; Urnovitz and Murphy, 1996; Wilkinson *et al.,* 1994). HERVs fall into several families according to sequence homologies, but the nomenclature of these families is not yet stabilized. Many of these proviruses are expressed in several organs, especially in the placenta and in tumor tissues, but most HERVs carry defective (i.e., noncoding and/or truncated) *gag, pol,* and *env* genes. However, a few HERV proteins were previously reported to be expressed under physiological conditions, particularly envelope proteins. The erv3/HERV-R envelope, for instance, is expressed in the syncytiotrophoblast layer of the placenta (Venables *et al.,* 1995). One provirus of the HERV-W family also encodes an envelope expressed in the placenta (Blond *et al.,* 1999). This HERV-W envelope is able to induce cell fusion in cells in culture upon interaction with the type D retrovirus receptor (Blond *et al.,* 2000). Besides a possible, but still hypothetical, fusogenic role *in vivo* in placental development (Mi *et al.,* 2000), HERV envelopes could also (i) protect against infection by closely related exogenous retrovirus via receptor interference (reviewed in Boeke and Stoye, 1997), and (ii) participate in the protection of the fetus from the mother's immune system via a domain (the immunosuppressive, or ISU, domain) located in the transmembrane subunit of most retroviral envelopes and known to inhibit immune effector functions (Cianciolo *et al.,* 1985; reviewed in Harris, 1998).

The HERV-H family is one of the most represented among human endogenous retroviruses, with approximately 1000 elements per haploid genome (Mager and Henthorn, 1984). HERV-H elements first entered the primate genome >40 million years ago since they are present in the genome of New World monkeys, but the majority of them result from a later expansion, some 30–35 million years ago (Anderssen *et al.,* 1997; Mager and Freeman, 1995). Expression of HERV-H elements was previously detected at a high level in normal placenta (Wilkinson *et al.,* 1990) and at lower levels in lung (Lindeskog *et al.,* 1993) and peripheral blood mononuclear cells (Kelleher *et al.,* 1996; Lindeskog *et al.,* 1993; Medstrand *et al.,* 1992). Several tumor cell lines, mainly teratocarcinoma, bladder carcinomas, testicular tumors, and lung tumors, also express significant levels of HERV-H transcripts (Hirose *et al.,* 1993; Wilkinson *et al.,*

[1] To whom correspondence and reprint requests should be addressed. Fax: 33-1.42.11.53.42. E-mail: heidmann@igr.fr.

0042-6822/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.

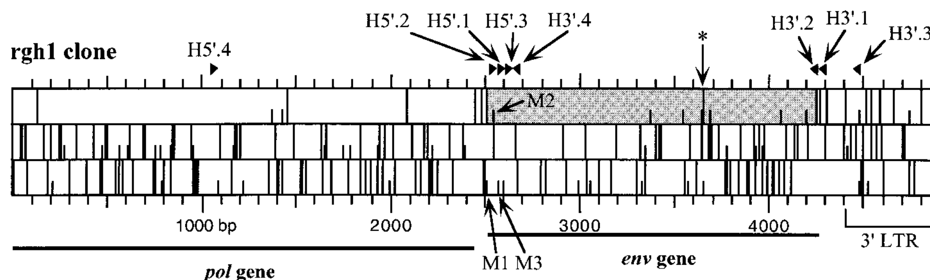


FIG. 1. Structure of the archetypal HERV-H rgh1 clone and position of the primers used in the present study. The ORFs for the three phases are figured, with complete vertical bars representing stop codons and short vertical bars representing methionines. Arrows point to the methionines referred to in the text as M1, M2, M3, to the stop codon interrupting the *env* ORF (asterisk), and to the primers (arrowheads).

1990). Most of the HERV-H proviruses are partially deleted, in that they are only 5.8 kb long, but a small subset (about 10%) is structurally intact, having full-length *gag, pol,* and *env* domains (Hirose *et al.,* 1993; Wilkinson *et al.,* 1993). The prototype full-length sequence, rgh2, is 8.7 kb long but its *gag, pol,* and *env* genes are interrupted by several stop codons. The incomplete rgh1 cloned provirus contained the first almost open envelope reading frame, interrupted by a unique stop codon located at the second third of the ORF (near the SU/TM junction) (Hirose *et al.,* 1993). In the course of this study, a first HERV-H *env* gene with a complete open reading frame was described (Lindeskog *et al.,* 1999). Considering that approximately 100 HERV-H envelope genes are present in the genome (Hirose *et al.,* 1993), we tried to isolate, using a protein truncation test (PTT), all the HERV-H envelope genes coding for complete retroviral envelopes. Here we show that only three HERV-H proviruses have the coding capacities for large envelopes encompassing the ISU domain (which is located in the very 3′ part of the ORF). These proviruses were characterized in humans and in the course of primate evolution.

RESULTS

Identification of HERV-H envelopes with large open reading frames by PTT on total genomic DNA

To select 5′-primers appropriate for PCR amplification of the HERV-H envelope gene copies and further *in vitro* transcription and translation by PTT, we first amplified a 1.2-kb HERV-H domain encompassing the 3′ end of *pol* and the 5′ end of *env* using the 5′-primer H5′.4 and the 3′-primer H3′.4, devised using the previously characterized rgh1 sequence as a reference (Fig. 1). The amplified fragments were cloned and 10 clones sequenced. At variance with the rgh1 clone in which the first methionine of the envelope ORF is encoded by nucleotides (nts) 2545–2547 (M2 in Fig. 1), the first Met was encoded by nts 2502–2504 (M1 in Fig. 1) in the 10 clones, because of the absence in all cases of nt 2538 (thus resulting in in-phase M1 and M2).

We then designed three different 5′-primers, one located between M1 and M2 (H5′.2) and the two others downstream of M2 (H5′.1 and H5′.3), and three different 3′-primers, two located between the first stop codon and the beginning of the 3′ LTR (H3′.1, H3′.2) and the other located in the 3′ LTR (H3′.3) (Fig. 1). The 5′-primers contained a T7-spacer-Kozak sequence, allowing direct transcription and translation of the PCR product (PTT). PCRs were performed on total human genomic DNA using seven combinations of primers, and the amplification products were subjected to *in vitro* transcription/translation. Depending on the primers used, different translation products were obtained (Table 1; Fig. 2A, lane 1; Fig. 2B, lane 1). Considering the length of the rgh1 open reading frame (without the central stop codon), the HERV-H envelope should have a size of 62 kDa; actually, three large translation products, of approximately 62, 60, and 59 kDa, were obtained by PTT. Other proteins, ≤42 kDa, were also obtained with some of the primers (see Table 1 and Fig. 2).

Chromosomal assignment of the corresponding envelope genes

Considering the expected number of *env*-containing HERV-H elements, the envelope PCR amplification product obtained from total genomic DNA should contain approximately 100 different molecular species. We therefore performed a PTT on isolated chromosomes, which on average should contain four HERV-H copies, to facilitate the cloning and characterization of the PCR products of interest. PTT was performed on a panel of monochromosomal hybrids, using two primer pairs: H5′.2–H3′.1 and H5′.2–H3′.3. The assay disclosed that the 62- and 59-kDa coding genes were located on chromosome 2 and the 60-kDa coding gene on chromosome 3 (Figs. 2B and 2C). Translation products of intermediate size were also observed. For example, proteins of about 42 kDa were encoded by chromosomes 1, 14, and 19 (Table 1, Fig. 2D, and data not shown), and of about 38 kDa by chromosome 12 (Fig. 2B, lane 3). As can be seen in Table 1, some translation products could be detected on isolated chromosomes but not on total genomic DNA, even

TABLE 1

Translational Products Obtained by PTT with HERV-H env Primers on Total Human Genomic DNA and on Monochromosomal Hybrids

DNA	PCR primer pair	p62[a]	p60[a]	p59[a]	p42[a]	p38[a]
Total genomic DNA	H5′.1–H3′.2	+	+	+	+	+
	H5′.1–H3′.3	−	−	+	−	−
	H5′.2–H3′.1	+	−	+	−	−
	H5′.2–H3′.2	+	+	+	+	+
	H5′.2–H3′.3	+	−	+	−	+
	H5′.3–H3′.1	+	+	+	−	+
	H5′.3–H3′.2	−	+	+	−	+
Chromosome 2	H5′.2–H3′.1	−	−	+	−	−
	H5′.2–H3′.3	+	−	+	−	−
Chromosome 3	H5′.2–H3′.1	−	+	−	−	−
Chromosome 12	H5′.2–H3′.3	−	−	−	−	+
Chromosome 19	H5′.2–H3′.1	−	−	−	+	−

[a] As 5′.1 and 5′.3 primers are located downstream of 5′.2 and yield shorter translational products, the sizes of the indicated translational products refer to the 5′.2-primed products (i.e., have been corrected when the two former primers are used).

using the same pair of primers: this was the case for p60 (detected with the H5′.2–H3′.1 primer pair on chromosome 3 but not on total genomic DNA) and for p42 (detected with the H5′.2–H3′.1 primer pair on chromosome 19 but not on total genomic DNA), most probably resulting from the fact that amplification of one definite copy among a few copies should be easier than among a hundred.

PCR products from chromosomes 2 and 3 were then cloned and a series of plasmids were submitted to PTT to analyze their coding potential. Plasmids containing genes coding proteins of the corresponding sizes (i.e., 59, 60, and 62 kDa) were obtained, and some of them were sequenced. In that these envelope genes were later resequenced directly from the BAC clones containing the complete corresponding proviruses, they are described below.

Cloning of the corresponding proviruses and precise chromosomal localization

Considering the high number of HERV-H envelope genes in the human genome and their high sequence homology, hybridization or PCR strategies as methods to clone the three envelope genes defined above were ruled out. In collaboration with the Centre d'Etude du Polymorphisme Humain (CEPH), we made use of a BAC human library arranged in pools and superpools, that we screened by combined PCRs and PTTs. The CEPH human BAC library is organized in 35 superpools, each containing 6144 BAC clones, in which the human genome is represented six times. We first applied PTT on the 35 superpools using the H5′.2–H3′.3 primer pair. The 62-, 59-, and 38-kDa envelope proteins were obtained in five, seven, and five superpools, respectively, out of 35.

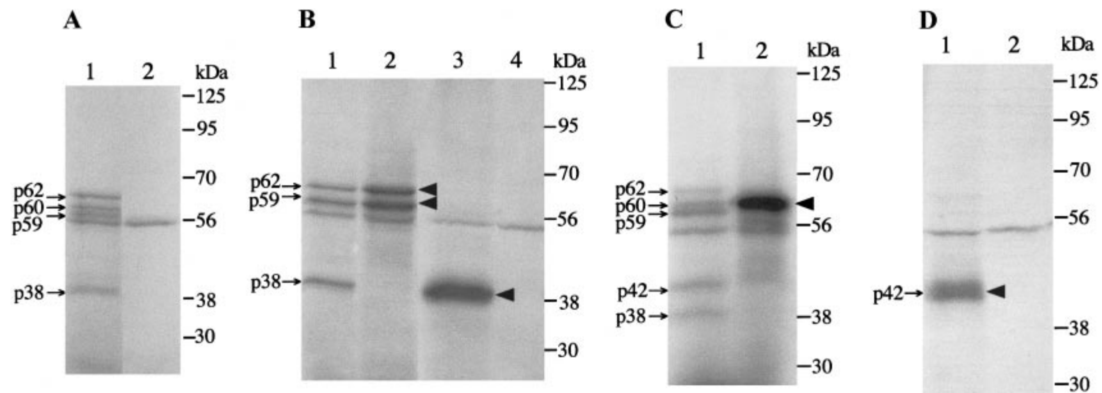


FIG. 2. Protein truncation test (PTT) of HERV-H envelopes in human total genomic DNA or DNA from monochromosomal somatic hybrids, using (A) the H5′.3–H3′.1 primer pair on total genomic DNA (lane 1) and no template (lane 2); (B) the H5′.2–H3′.3 primer pair on total genomic DNA (lane 1), chromosome 2 somatic hybrid DNA (lane 2), chromosome 12 somatic hybrid DNA (lane 3) and no template (lane 4); (C) the H5′.2–H3′.2 primer pair on total genomic DNA (lane 1) and chromosome 3 somatic hybrid DNA (lane 2); (D) the H5′.2–H3′.1 primer pair on chromosome 3 somatic hybrid DNA (lane 1) and no template (lane 2). Arrows indicate specific translation products with their corresponding size, and arrowheads point to those obtained from single chromosomes.

Superpool number 4 happened to encode all three proteins and was therefore selected for further analysis. The next screening was performed by PCR with the same pair of primers as the one used for PTT, using a matrix approach to minimize the number of PCRs. This led to the isolation of one candidate BAC. A final PTT with this BAC yielded a translation product of 62 kDa. Superpool number 15, which was found to encode the 62- and 59-kDa proteins, was further assayed to isolate the 59-kDa coding envelope. PCR screening (again with the H5′.2–H3′.3 primer pair) yielded three candidate BACs, which were tested by PTT and two of which encoded the 59-kDa protein. A new primer pair, H5′.3–H3′.2, was used to screen the superpools by PTT, to isolate a BAC containing the provirus encoding the 60-kDa protein. The 60- and 59-kDa proteins were each obtained in six superpools out of the 35. Superpool number 30 encoded both 60- and 59-kDa proteins and was chosen to be PCR-screened with the H5′.3–H3′.2 primer pair, which led to two candidate BACs, one of which yielded a 60-kDa protein by PTT and the other, a 59-kDa protein. Sequencing of a 500-bp fragment from each of the three p59-encoding BACs isolated by the two screenings revealed that they contained the same gene.

To further define the localization of the three proviruses, chromosome metaphases were analyzed by fluorescence *in situ* hybridization (FISH) using the corresponding BACs (B231E12, H734E12, and B916F3) as probes. Each BAC gave a single hybridization signal per haploid genome, on the same chromosomes as those identified by PTT in the monochromosomal hybrids panel experiment reported above. The precise localization is 2q24.3 for HERV-H/p62, 2q23 for HERV-H/p59, and 3q26 for HERV-H/p60 (data not shown).

Analysis of the three proviral structures in humans

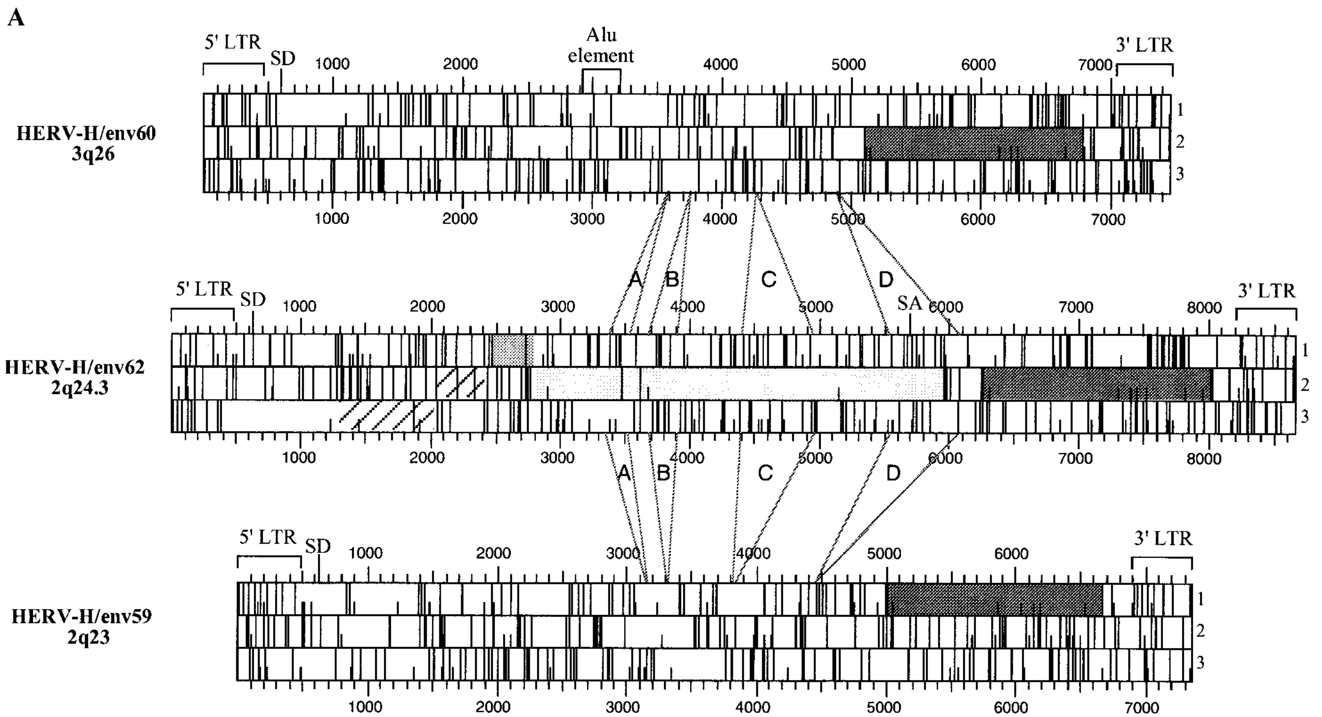
Using primers scattered along the rgh2 HERV-H prototype provirus sequence, the DNA of the three BACs (each encoding one of the three proteins) was directly sequenced. When a sequence could not be obtained with a primer on a given BAC, a new sequencing primer was designed, and so on, until the complete sequence of the three proviruses was obtained. Figure 3A shows the ORF maps of the three proviruses that we named HERV-H/env62, HERV-H/env60, and HERV-H/env59, referring to the sizes of the putative envelope proteins. The presence of a 5-bp cellular sequence duplication (imperfect for HERV-H/env62; see Fig. 4) bordering the three proviruses indicates that they integrated through retroviral transposition. The LTRs vary in length from 416 bp (for the 5′ HERV-H/env60 LTR) to 497 bp (for the 5′ HERV-H/env59 LTR). Two out of the three previously characterized subtypes of HERV-H LTRs (Goodchild *et al.,* 1993; Mager, 1989) are represented here, with HERV-H/env62 and HERV-H/env59 LTRs belonging to type I LTRs, and HERV-H/env60 LTRs to type II. A histidine tRNA primer binding site is found 2 nts downstream of the 5′ LTR and a small purine stretch (12 residues interrupted by one pyrimidine) is present just upstream of the 3′ LTRs for the three proviruses.

Among the three proviruses, HERV-H/env62 is the one that contains the longest ORF in the expected *gag* gene. A BLAST computer search against a viral division of GenBank ascertained the *gag* origin of the domain between nt 1300 and nt 2400 (according to the nucleotide numbering in Fig. 3A) for this HERV-H element, upon alignment with HTLV-1, Mo-MLV, and BaEV gag proteins. A sequence possibly corresponding to the major homology region (located at the carboxyl-terminus of the capsid subdomain of *gag* in most retroviruses) (Craven *et al.,* 1995) can be tentatively positioned around nt 1900 upon alignment with the Mo-MLV and BaEV gag proteins, but did not display the three conserved residues of this domain. Similarly, two characteristic motifs of the nucleocapsid subdomain of *gag* ($CX_2CX_4HX_4C$ domains) (Covey, 1986) can be identified around nt 2280 upon alignment with the HTLV-1 gag protein, but with an imperfect match.

The *pol* gene of HERV-H/env62 disclosed significant similarities with the Mo-MLV *pol* gene, which initiate at nt 2460 and stop at nt 5960. An amino acid homology analysis program [framesearch from the Genetics Computer Group (Madison, WI) package] revealed that the protease-encoding region is not in the same reading frame as the RT and the integrase. Four regions, designated A, B, C, and D and located in the *pol* region, were previously identified as present only in HERV-H full-length elements (Lindeskog and Blomberg, 1997; Wilkinson *et al.,* 1993). Recently, *env*-containing HERV-H elements were described, devoid of A, B, and C regions (Lindeskog *et al.,* 1999). Two out of the three proviruses described here (HERV-H/env60 and HERV-H/env59) lack all four regions (see Fig. 3A). The D region contains the splice acceptor sites used to generate *env* subgenomic mRNAs in the HERV-H family (Lindeskog and Blomberg, 1997). It is worth noting that an Alu element, 300 nt long, is found in the 5′ part of the HERV-H/env60 *pol* gene (see Fig. 3A).

The three identified envelope genes have, on average, 95% similarity in nucleotides and 90% in amino acids. All three genes display strong homology with the rgh1 *env* (94% in nt, on average; see Fig. 3B). As can also be seen in Fig. 3B, the first stop codon for the p62 *env* is the same as in the rgh1 clone, while the two smaller proteins use a common stop codon located 21 aa upstream. In addition, the p56 *env* harbors an 8-aa deletion in the 3′ half of the consensus retroviral immunosuppressive domain. The three envelope gene products described here disclose hydrophobicity profiles as well as other characteristic features of retroviral envelopes, i.e., a putative signal peptide located just downstream of the M2 methio-

A



B

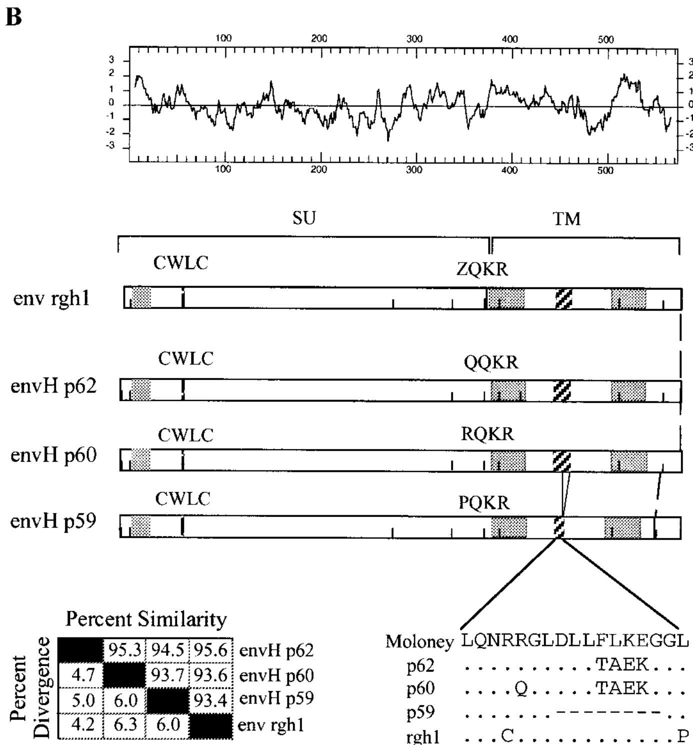


FIG. 3. Structure of the three HERV-H proviruses disclosing large open *env* reading frames and characterization of the corresponding putative proteins. (A) ORF map of the three HERV-H proviruses, with the LTRs and the splice donor and acceptor sites (SD, SA) indicated. The A, B, C, D regions present in HERV-H/env62 are deleted in HERV-H/env59 and HERV-H/env60, as observed in the majority of the HERV-H; the open envelope reading frames are in dark gray, the protease and the polymerase reading frames of HERV-H/env62 in intermediate and pale gray, respectively; the gag homology region of HERV-H/env62 is delineated with stripes. (B) Structure of the corresponding HERV-H envelopes and their expected functional domains in comparison with the rgh1 prototypic HERV-H envelope. The open rectangles delineate open reading frames, with the grayed domains corresponding to hydrophobic regions associated with the signal peptide, the fusion peptide, and the transmembrane region, respectively; the sequences of the CWLC motifs and the proteolytic cleavage sites (located just upstream of the fusion peptide) are indicated above the open reading frames; the putative immunosuppressive domains are represented with striped boxes and their sequences compared with that of Mo-MLV are given below, with the identical positions indicated with dots and gaps with dashes. A hydrophobicity profile of the putative rgh1 envelope (similar profiles were obtained for the three presently identified envelopes) and values of nucleic acid similarities between the four envelopes are given on top and bottom of the figure, respectively.

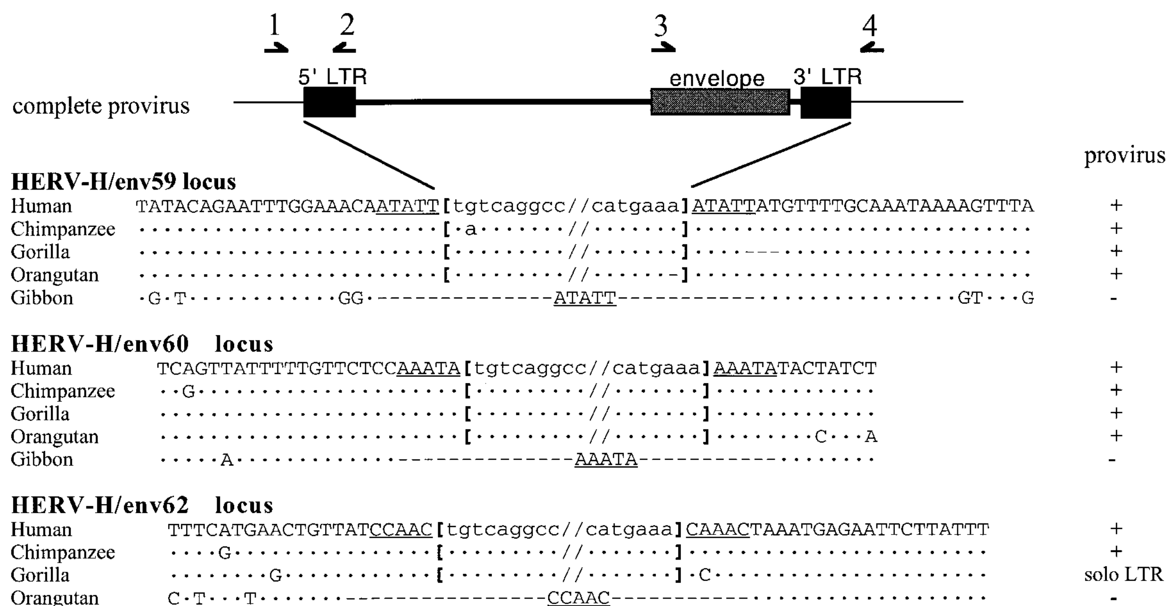


FIG. 4. Status and characterization of the loci associated with three env-encoding HERV-H proviruses among primates. The structure of the HERV-H proviruses together with the location of the primers used for the PCRs are schematized in the upper part of the figure. The sequences of the proviral boundaries in the indicated primate species are given in the lower part, with the proviral sequences, when present, in brackets, and the 5-bp target site duplications are underlined. On the right, the status of the three loci are schematized, with "+" to indicate a full-length provirus in the corresponding target site, and "−" for the empty site; only a solo LTR is present in Gorilla for the HERV-H/env62 locus.

nine, a CWLC motif (involved in the linkage between SU and TM subunits), a furin cleavage site at the junction of the two subunits (which does not fit the consensus R/KXR/KR, except for the p60 protein) followed by a hydrophobic fusion peptide, a putative immunosuppressive domain, and a hydrophobic membrane-spanning domain (see Fig. 3B). We compared the nucleotide sequence of the HERV-H complete envelope recently cloned by Lindeskog (HERV-H19) (Lindeskog *et al.,* 1999) with that of the three presently identified envelope genes, and found two single nucleotide differences with the HERV-H/env62 envelope, resulting in two amino acid substitutions. To exclude the existence of a fourth fully open HERV-H envelope gene, and to determine whether the HERV-H/env62 gene could be polymorphic in the human population, we sequenced the complete p62 envelope gene from five unrelated individuals. Sequence analysis revealed that these two nucleotide differences indeed corresponded to polymorphic sites (see Table 2).

Dating the three proviral insertions

Since retroviruses integrate randomly in the genome, the presence of a provirus at the same given locus among different species dates the integration event in the germline of their common ancestor. A classical PCR strategy was therefore used to characterize the distribution of the three above-identified proviruses among primates (Fig. 4). The presence of a provirus was assessed by PCR amplification (i) of its 5′ LTR and flanking DNA (amplification with primers 1 and 2) and (ii) of the region encompassing the envelope, the 3′ LTR and flanking DNA (amplification with primers 3 and 4). Absence of the provirus was assessed by amplification of the empty target site using the two flanking primers (amplification with primers 1–4), followed by sequencing of the PCR product. As can be inferred from Fig. 4, which provides the sequences of the junction between the target DNA and the provirus for several primates, the HERV-H/env62 provirus integrated in the germline of the common ancestor of gorillas, chimpanzees, and humans, after the split of the orangutan lineage. In the gorilla lineage, the presence of a solo LTR indicates a recombination event leading to excision of the internal proviral sequence. On the other hand, HERV-H/env60 and HERV-H/env59 proviruses are present in the orangutan genome but not in

TABLE 2

Identification and Distribution of Polymorphic Sites within the HERV-H/env62 Envelope Gene

Position[a]	Nucleotide	Amino acid change	No. of chromosomes[b]	Distribution
6553	G/T	Val/Leu	13	8/5
6762	T/A	Phe/Leu	13	8/5

[a] According to the nucleotide numbering of the HERV-H/env62 provirus deposited in GenBank with accession number AJ289709.

[b] Corresponding to diploid genomes (from five unrelated individuals) and haploid sequences (from a BAC, a monochromosomal hybrid PCR fragment, and the HERV-H19 sequence).

the gibbon genome, indicating that they integrated in the germline of the common ancestor of great apes.

A second method to date the integration of a provirus is to measure the sequence divergence between its 5′ and 3′ LTRs. Indeed, the two LTRs should be identical at the time of integration, as a result of the molecular mechanism of retroviral replication, and the percentage of divergence between the two LTRs is therefore an indication of the time elapsed since this initial event. In the present case, the percentages of divergence between the two LTRs were 4.3, 4.3, and 2.4% for HERV-H/env62, HERV-H/env60, and HERV-H/env59, respectively. The fact that the percentage is lower for HERV-H/env59 than for HERV-H/env62, despite the fact that the latter integrated more recently in the primate lineage, is indicative of a bias in the method and actually strongly suggests that homogenization events take place between LTRs. To further investigate this phenomenon, the 5′ and 3′ LTRs of the three proviruses in each primate species were sequenced and their structures compared. As illustrated in Fig. 5A, the number of internal repeats present in each LTR already reveals a structural diversity. This diversity exists for a given provirus among species, and also between the 5′ and 3′ LTRs of a given provirus, e.g., for HERV-H/env60 in humans and in gorillas, or HERV-H/env59 in humans.

Recently, Johnson and Coffin (1999) showed that HERVs can be used as phylogenetic markers by constructing trees with the LTRs of proviruses among primates. Assuming that the two LTRs are identical at the time of integration and are evolving independently, the repartition of 5′ and 3′ LTRs of a particular provirus in different species should display a separate clustering of the 5′ and the 3′ LTRs, with each cluster having the same topology as the standard tree of primate evolution. When we similarly built phylogenetic trees using the LTRs of the three proviruses among the primate species, we observed within each major branch (corresponding to each proviral integration) a strong deviation of the resulting trees from the predicted model, with very frequent clustering of the 5′ and 3′ LTRs from the same primate species (especially in the orangutan lineage) (Fig. 5B). This, again, is a strong hint for the occurrence of homogenization events between the two LTRs of the proviruses.

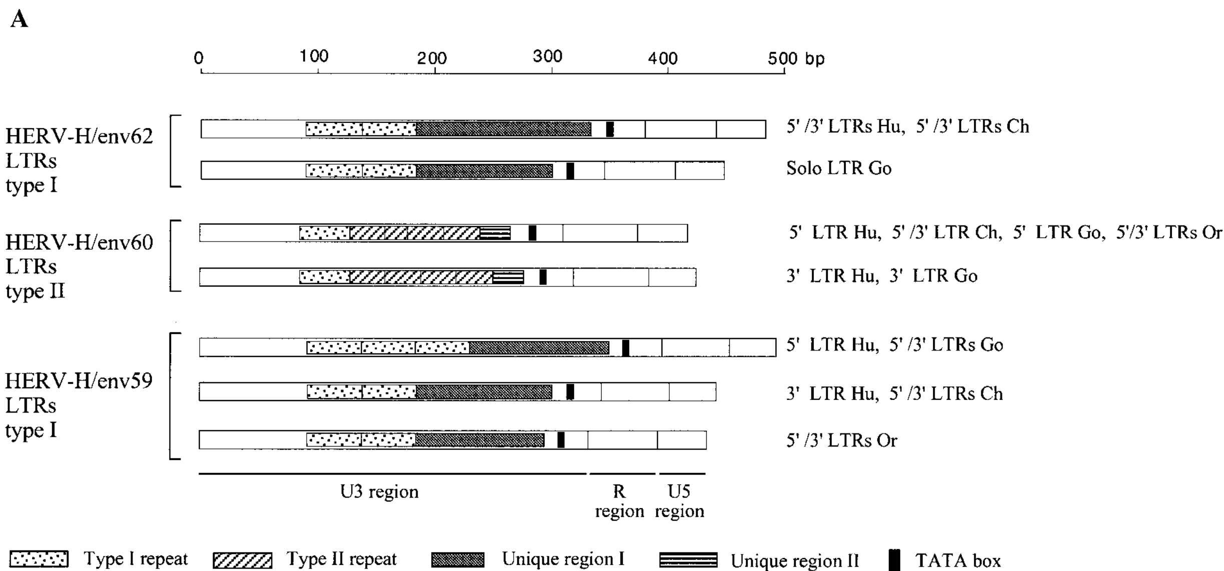
Analysis of the HERV-H envelope open reading frames in the primate genomes

To determine whether the three coding envelope genes identified in humans are also open in the primates carrying the corresponding proviruses, we performed a PTT specific for each of the three envelopes on genomic DNAs, using the H5′.2 forward primer and the reverse flanking primers specific for each provirus. As shown in Fig. 6, HERV-H *env* genes corresponding to HERV-H/env62, HERV-H/env60, and HERV-H/env59 encode smaller proteins in primates, except for the p59 envelope in chimpanzees for which the ORF is as long as that in humans. In two cases (HERV-H/env60 in chimpanzees and HERV-H/env59 in gorillas), no protein could be detected by PTT. Expectedly, sequencing of the two corresponding envelope PCR products revealed premature stop codons, leading to putative proteins of 10 kDa for HERV-H/env60 in chimpanzees and of 6 kDa for HERV-H/env59 in gorillas. To determine whether other coding-competent HERV-H envelope genes are present in the primate genomes apart from the three presently described loci, we performed a "general HERV-H envelope PTT" on genomic DNAs from chimpanzees, gorillas, and orangutans, using two internal *env* primer pairs (H5′.2–H3′.2 and H5′.3–H3′.1). Several translation products were obtained (among which, as expected, were the proteins associated with the three specific loci), ranging from 60 to 25 kDa. Interestingly, the chimpanzees' and the gorillas' genomic DNA revealed a 60-kDa HERV-H envelope, and the orangutans', a 59-kDa HERV-H envelope (data not shown). These envelope genes are necessarily associated with a fourth HERV-H locus, since the three proviruses characterized above encode smaller proteins in these primates (see Fig. 6).

DISCUSSION

Downsizing of a multigene family

In this report, we show that among one of the largest families of HERVs, the HERV-H family, only three elements maintained a large open envelope reading frame. We showed that the HERV-H19 envelope gene cloned recently by Lindeskog *et al.* (1999), using a hybridization screening method with an *env* fragment on a genomic library, is an allelic variant from HERV-H/env62. Several arguments allow us to assert that no other completely coding HERV-H envelope gene exists in the human genome. First, we used a redundant series of primers for the screening of the human genome for open *env* reading frames by the protein truncation test (PTT). PTTs using total human genomic DNA and monochromosomal hybrids actually led to the same results, that is, occurrence of only three large proteins of 59, 60, and 62 kDa, one of which is encoded by chromosome 3 and the two others, by chromosome 2. Moreover, when the genomic loci were cloned from the BAC library, using PTT as a screening procedure, the frequency of occurrence for each protein was fully compatible with a unique representation in the genome. In addition, sequences of the BACs coding for these proteins were almost identical to the sequences of the PCR products obtained from the monochromosomal hybrids, and the localization of the BACs by FISH was also in agreement with the results from the PTT on monochromosomal hybrids. Finally, a computer search in databases recently performed (on June 29, 2000) did not reveal any other envelope (with the



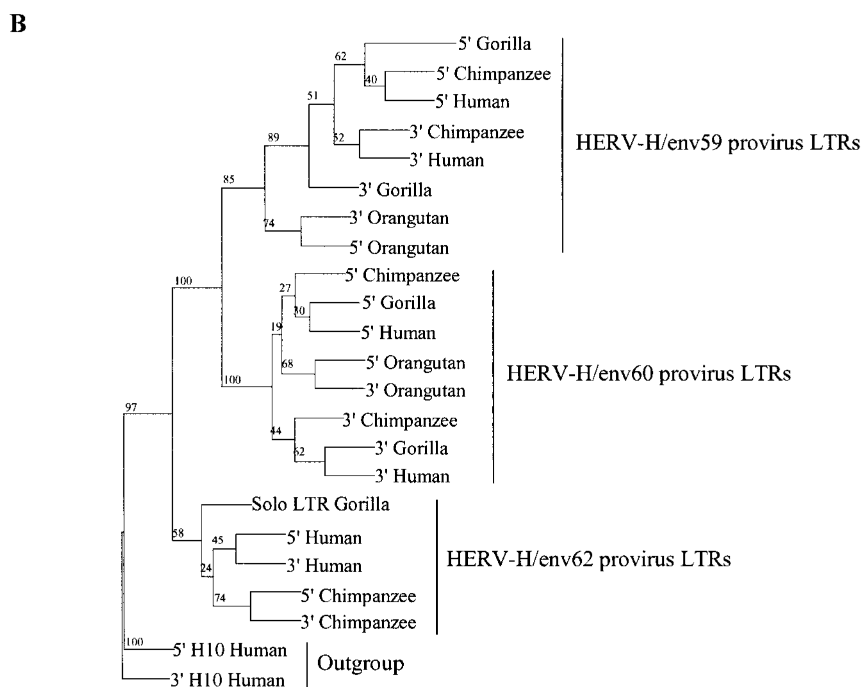


FIG. 5. Structure of the LTRs of the three env-encoding HERV-H among primates, and phylogenetic analysis. (A) Schematic representation of the LTRs of the three proviruses among primates, grouped according to their subtype as defined from the nature of the repeats in the U3 region. Hu, human; Ch, chimpanzee; Go, gorilla; Or, orangutan. (B) Phylogenetic tree of the LTRs of the three proviruses in primates (neighbor-joining tree constructed with the U3-R region). Parsimony methods yielded similar results (data not shown). Numbers indicate bootstrap values for 100 replicates, and branch lengths indicate the degree of divergence between sequences. The outgroup used in this study contains the HERV-H10 LTRs (accession numbers AF108838 and AF108841).

exception of HERV-H19; see above), thus confirming the exhaustiveness of the present investigation. In this respect it is worth noting that, although the Human Genome Program should have released >82% of the human genome, the p59 and p60 encoding genes were still not found at that time.

Two other HERV families were tentatively screened for complete envelope genes, the HERV-K (Mayer *et al.,* 1997; Tönjes *et al.,* 1999) and the HERV-W families (Voisset *et al.,* 2000). Both families comprise approximately 30 *env*-containing proviruses, and both entered the primate genome 25–30 million years ago. The HERV-K family was shown by both PTT analysis and screening of a genomic library to contain two fully coding envelopes. The HERV-W was screened by a PTT approach, and was shown to contain only one complete envelope. Because of the screening procedures (only one pair of primers was used for PTT in these studies), a definite conclusion

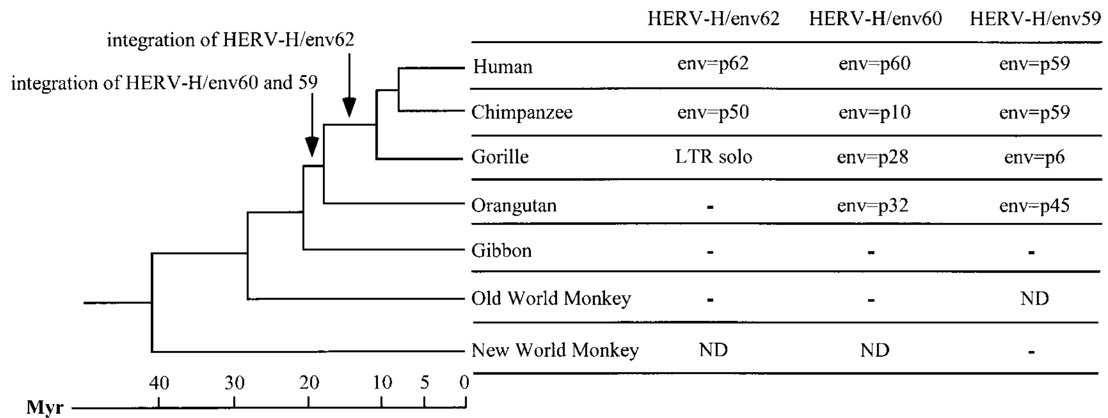


FIG. 6. Coding capacity of the envelope genes of the three HERV-H in primates. In the left part of the figure a scheme of primate evolution is presented, with the integration times for the three proviruses. In the right part of the figure, the length of the open envelope reading frames in the indicated primates is given for each of the HERV-H locus mentioned. Sizes in kDa were derived from the PTT, and by direct sequencing of the PCR products when no translational product could be detected by PTT (products ≤ 10 kDa). —, absent provirus; ND, status of the locus not determined (no amplification obtained).

cannot be derived for these two families. Yet, the proportion of coding versus noncoding envelope genes seems to be comparable within the three families (three out of 100 for HERV-H, two out of 30 for HERV-K, and one out of 30 for HERV-W).

Evolution of integrated retroviral genomes

One interesting outcome of the present investigation of the three HERV-H proviruses among primates is the demonstration that these sequences are submitted to a series of alterations, some of which are possibly specific for proviral structures. Actually, we could demonstrate that dating of the insertion of the HERV-H proviruses described here could not rely on the degree of divergence between their 5′ and 3′ LTRs, as homogenization phenomena between both LTRs were observed for each of the proviruses. This phenomenon does not necessarily concern all HERV-H elements since, in the Johnson and Coffin (1999) study, the RTVL-Ha provirus LTR tree is totally congruent to the expected topology. As discussed in their study, homogenization of both LTRs of a provirus probably occurred by intrachromosomal gene conversion (i.e., transfer of a portion of sequence by homologous recombination). The existence of a solo LTR for the HERV-H/env62 locus in the gorilla genome is an additional evidence of recombination events in endogenous retroviral genomes. Another LTR recombinational phenomenon seems to be superimposed on this homogenization pattern, which to our knowledge was never described before. As shown in Fig. 6, the structure of the U3 region of a given provirus varies between species, and even between the 5′ and 3′ LTR within a particular species. This variation concerns short-sequence (<30 bp) duplications or deletions, which probably arose through a slipped-strand mispairing mechanism during cellular DNA replication (Levinson and Gutman, 1987).

A role for coding endogenous retroviral genes?

Only three HERV-H out of 100 *env*-containing proviruses still possess an open envelope reading frame. There are *a priori* two possible reasons why open reading frames of transposable elements should be maintained over such a long period: either because the ORFs are of benefit to the host or because the ORFs are essential for the parasite. Numerous roles beneficial to the host have been hypothesized, including protection against infection and basic roles in placenta formation or placental local immunosuppression (reviewed in Harris, 1998). Although such roles cannot be excluded, one has to acknowledge that until now only protection against infection has been convincingly demonstrated, with the unraveling of the molecular nature of the Fv loci in mice (reviewed in Boeke and Stoye, 1997). The alternative interpretation simply states that functional proteins are maintained only by chance, as a result of a favorable balance between the genetic drift of individual integrated elements and the rate of replication of still functional ones making "safety" copies within the family. This interpretation, already highlighted in our previous analysis of an ancient family of mammalian ERVs (Bénit *et al.,* 1999), does not impose any selective pressure from the host for the maintenance of open reading frames, and simply accounts for the fact that some HERV families are most probably already dead ends that are no longer functional as a family, whereas other are surviving precisely because they can still retrotranspose. This replicative function is not necessarily associated with definite copies and/or loci (taking into account that some retroviral functions can be complemented *in trans* among dispersed elements), consistent with our present observation that among primates open envelopes are not carried by the same integrated proviruses.

Along this line, the situation of the *env* gene is paradoxical. Indeed, this gene, which is initially required for the entry of the virus into the host, in principle should not be necessary for the intracellular transposition of the integrated provirus, which requires only the *gag* and *pol* genes for its replication (Tchenio and Heidmann, 1991, 1992). Although it was previously shown in the case of the mouse that reinfection of the germline by viral particles of endogenous origin might be responsible for provirus expansion (Lock *et al.,* 1988; Panthier *et al.,* 1988), this is not the case for several Drosophila endogenous retroviruses, which have a still fully coding *env* gene but do not require this gene for their germline transpositional activity (e.g., gypsy) (Chalvet *et al.,* 1999). Accordingly, it seems finally plausible that endogenous *env* genes that encode full-length proteins, as presently observed, may have actually lost their initial virus-associated function, without having necessarily and already gained a new function of benefit to the host.

Perspectives

Besides possible roles at the time-scale of species evolution, it has to be kept in mind that endogenous retroviruses disclose impressive similarities with exogenous infectious retroviruses, from which they most probably derive. Therefore, it remains possible that endogenous retroviruses have conserved some features of exogenous retroviruses that could be responsible for pathogenic "somatic" effects. In this respect, we previously demonstrated that the immunosuppressive activity of the Mo-MLV envelope gene could be involved in a tumorigenic process, at least in an *in vivo* assay involving the engrafting of tumor cells to immunocompetent mice that can normally reject these cells (Mangeney and Heidmann, 1998). Since the HERV-H *env* contains a sequence most closely related to the so-called immunosuppressive domain of Mo-MLV, also shared by other infectious retroviruses (Cianciolo *et al.,* 1985; Mangeney and Heidmann, 1998), we are now testing its possible immunosuppressive effect in the above-mentioned *in vivo* assay and already have strong hints that, at least, the longest HERV-H *env* (envH p62) has the capacity to allow tumor cells to escape immune rejection in immunocompetent mice (unpublished results). Interestingly, HERV-H elements are induced in several tumor tissues and cell lines (Hirose *et al.,* 1993; Wilkinson *et al.,* 1990), but it is still not known whether this induction involves one of the cloned HERV-H elements. Answering this question will require a refined analysis of the pattern of expression of the identified proteins via the development of antibodies, allowing the screening of human tissues for the expression of an HERV-H envelope.

In conclusion, this study provides *de facto* the first exhaustive, yet still limited, list of candidate genes for the possible involvement of HERV-H in a human physiopathology. The identified genes together with their chromosomal positioning should allow further studies of their function via classical genetic approaches [identification of susceptibility loci, search for polymorphisms among the human population as previously performed for the erv3 locus (de Parseval *et al.,* 1998)], whereas the limited number of unraveled coding sequences downsizes an *a priori* insoluble multigenic analysis to a simpler one.

MATERIALS AND METHODS

DNA samples

Human DNAs were extracted from peripheral blood leukocytes from healthy donors. The monochromosomal somatic hybrids panel was obtained from Coriell Institute (Camden, NJ). The Human BAC library was from the Centre d'Etude du Polymorphisme Humain (CEPH, Paris, France). Chimpanzee DNA was extracted from peripheral blood leukocytes from healthy animals provided by Françoise Barré-Sinoussi (Institut Pasteur). Gorilla (*Gorilla Gorilla*), orangutan (*Pongo pygmaeus*), and gibbon (*Hylobates Lar Moloch*) DNAs were extracted from primary fibroblast cells provided by Florence Richard (Institut Curie). Rhesus macaque (*Macaca cynomolgus*) DNA was prepared from tissue donated by Guy Germain (INRA, Jouy en Josas, France).

PCR and sequencing of PCR products

Primers used to amplify the 5′ part of the envelope were as follows: H5′.4 (CCCACACAAGGCAAATGGTTC), H3′.4 (CCCATTCTCTCCTAGCCGCTTCT). PCRs were carried out for 35 cycles (1 min at 94°C, 1 min at 60°C, 4 min at 72°C), using 50 ng of genomic DNA, 50 pmol of each PCR primer, 1.5 mM of MgOAc, and 0.2 mM of each dNTP, in 50 μl PCR reaction buffer (Tth polymerase and buffer; Perkin–Elmer, Palo Alto, CA). Resulting fragments were T/A cloned in pGEM-T easy (Promega, Madison, WI) and 200 ng of plasmid sequenced with the H3′.4 primer using the Applied Biosystems *Taq* DyeDeoxy Terminator Cycle Sequencing kit with an ABI 373A sequencer (Perkin–Elmer).

Primers for the amplification of the full-length envelope genes of the HERV-H family were as follows: three forward primers containing the T7 promoter and Kozak ATG, H5′.1 (GCTAATACGACTCACTATAGGAACAGACCACCATGCACCCTCTACCTCTCCCCAG), H5′.2 (GCTAATACGACTCACTATAGGAACAGACCACCATGAGGGCACCCTCCAATACTTC), and H5′.3 (GCTAATACGACTCACTATAGGAACAGACCACCATGCACCACAGTATCAACCTTAC); and three reverse primers, H3′.1 (CTCAGGGCTGCTTCAAGCGG), H3′.2 (GATTAAGCTGAAGGGAGATC), and H3′.3 (CATCAG/TTTAAGGT/CG/AGGGCAGGG). Amplification products obtained from the monochromosomal mapping panel were T/A cloned in pGEM-T easy (Promega). Plasmids of interest (200 ng) were sequenced using the following

primer set: H5′.2, HS5 (CCCTACACCTCCGAACTTCC), HS6 (CTCCGAAGCCCAACTACACAC), HS4 (GCAGTTACCCCATCAGTCCC), and H3′.1.

Primers used to amplify the envelope gene of HERV-H/env59, HERV-H/env60, and HERV-H/env62 were H5′.2 as forward primer and, respectively, fIR59 (ACCCCATGTTCTAGTCTTCC), fIR60 (AGCAATAGTTTGTTAAATTC), and fIR62 (GTGAGGGAAAGAATATTAGG) as reverse primers.

Primers used to amplify LTRs from primates and the proviral cellular target sequences were fIF59 (GAAAAGTAAAGAAAGAAGCC), fIF60 (GATTTCCTTGTATTCTGTCC), fIF62 (CTTAAGTCAATAATGCAGTG), fIR59, fIR60, fIR62, LTR5R (GAAGAGACCACCAAACAGG), and LTR3F (CCACCCCTAATACCGCTTGA). PCR products were gel-purified and sequenced directly with the primers used for amplification.

Protein truncation test (PTT)

T7 promoter-containing PCR products were ethanol precipitated and 200 ng of the amplification products were used in the TNT Coupled Reticulocyte Lysate System (Promega), according to the manufacturer's instructions, with [^{3}H]methionine (ICN, Irvine, CA) for protein labeling. After electrophoresis of the translation products, the SDS–polyacrylamide gels were impregnated in En^{3}hance (Dupont de Nemours, Geneva, Switzerland), rinsed with water, dried, and autoradiographed.

Sequencing of the BACs and computer sequence analyses

BAC DNAs were prepared using a Qiagen plasmid maxi kit (Qiagen, Chatsworth, CA) and a protocol modified following the instructions of the manufacturer. They were sequenced directly, following a procedure developed by V. Lazar at the Institut Gustave Roussy.

Sequences were analyzed using software provided by the Genetics Computer Group and homology searches in the databases were conducted using the BLAST server at NCBI. Alignments were performed with the CLUSTALW multialignment program.

Accession numbers

The HERV-H proviruses and the primate LTRs of the three proviruses were deposited in GenBank with accession numbers AJ289709, AJ289710, AJ289711, AJ291349, AJ291350, AJ291351, AJ291352, AJ291353, AJ291354, AJ291355, AJ291356, AJ291357, AJ291358, AJ291359, AJ291360, AJ291361, AJ291362, and AJ291363.

ACKNOWLEDGMENTS

We are very grateful to Vladimir Lazar for the sequencing of the proviruses in the BACs, to Hera Der Sarkissian for the localization of the BACs by FISH, to Laurence Bénit for assistance with nucleic acid analyses, and to Marianne Mangeney for helpful discussions. This work was supported by the CNRS and by grants from the Ligue Nationale contre le Cancer ("Equipe Labellisée") and by a fellowship from the Fondation pour la Recherche Médicale (to N.P.).

REFERENCES

Anderssen, S., Sjottem, E., Svineng, G., and Johansen, T. (1997). Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology* **234,** 14–30.

Bénit, L., Lallemand, J.-B., Casella, J.-F., Philippe, H., and Heidmann, T. (1999). ERV-L elements: A family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* **73,** 3301–3308.

Blond, J.-L., Besème, F., Duret, L., Bouton, O., Bedin, F., Perron, H., Mandrand, B., and Mallet, F. (1999). Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family. *J. Virol.* **73,** 1175–1185.

Blond, J.-L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., Mandrand, B., Mallet, F., and Cosset, F.-L. (2000). An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta, and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* **74,** 3321–3329.

Boeke, J. D., and Stoye, J. P. (1997). Retrotransposons, endogenous retroviruses, and the evolution of retroelements. *In* "Retroviruses" (J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds.), pp. 343–436. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Chalvet, F., Teysset, L., Terzian, C., Prud'homme, N., Santamaria, P., Bucheton, A., and Pélisson, A. (1999). Proviral amplification of the Gypsy endogenous retrovirus of Drosophila melanogaster involves env-independent invasion of the female germline. *EMBO J.* **18,** 2659–2669.

Cianciolo, G. J., Copeland, T., Orozlan, S., and Snyderman, R. (1985). Inhibition of lymphocyte proliferation by a synthetic peptide homologous to retroviral envelope protein. *Science* **230,** 453–455.

Covey, S. N. (1986). Amino acid sequence homology in gag region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* **14,** 623–633.

Craven, R. C., Leure-Du-Pree, A. E., Weldon, R. A. J., and Wills, J. W. (1995). Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. *J. Virol.* **69,** 4213–4227.

De Parseval, N., and Heidmann, T. (1998). Physiological knock-out of the envelope gene of the single copy ERV-3 human endogenous retrovirus in a fraction of the Caucasian population. *J. Virol.* **72,** 3442–3445.

Goodchild, N. L., Wilkinson, D. A., and Mager, D. L. (1993). Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* **196,** 778–788.

Harris, J. R. (1998). Placental endogenous retrovirus (ERV): Structural, functional, and evolutionary significance. *BioEssays* **20,** 307–316.

Hirose, Y., Takamatsu, M., and Harada, F. (1993). Presence of *env* genes in members of the RTLV-H family of human endogenous retrovirus-like elements. *Virology* **192,** 52–61.

Johnson, W. E., and Coffin, J. M. (1999). Constructing primate phylogenies from ancient retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **96,** 10254–10260.

Kelleher, C. A., Wilkinson, D. A., Freeman, J. D., Mager, D. L., and Gelfand, E. W. (1996). Expression of novel-transposon-containing mRNAs in human T cells. *J. Gen. Virol.* **77,** 1101–1110.

Levinson, G., and Gutman, G. A. (1987). Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4,** 203–221.

Lindeskog, M., and Blomberg, J. (1997). Spliced human endogenous retroviral HERV-H env transcripts in T-cell leukaemia cell line and normal leukocytes: Alternative splicing pattern of HERV-H transcripts. *J. Gen. Virol.* **78,** 2575–2585.

Lindeskog, M., Mager, D., and Blomberg, J. (1999). Isolation of a human

endogenous retroviral HERV-H element with an open *env* reading frame. *Virology* **258,** 441–450.

Lindeskog, M., Medstrand, P., and Blomberg, J. (1993). Sequence variation of human endogenous retrovirus ERV9-related elements in an *env* region corresponding to an immunosuppressive peptide: Transcription in normal and neoplastic cells. *J. Virol.* **67,** 1122–1126.

Lock, L. F., Keshet, E., Gilbert, D. J., Jenkins, N. A., and Copeland, N. G. (1988). Studies of the mechanism of spontaneous germline ecotropic provirus acquisition in mice. *EMBO J.* **7,** 4169–4177.

Löwer, R., Löwer, J., and Kurth, R. (1996). The viruses in all of us: Characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl. Acad. Sci. USA* **93,** 5177–5184.

Mager, D. L. (1989). Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family RTVL-H. *Virology* **173,** 591–599.

Mager, D. L., and Freeman, J. D. (1995). HERV-H endogenous retroviruses: Presence in the New World branch but amplification in the Old World primate lineage. *Virology* **213,** 395–404.

Mager, D. L., and Henthorn, P. S. (1984). Identification of a retrovirus-like repetitive element in human DNA. *Proc. Natl. Acad. Sci. USA* **81,** 7510–7514.

Mangeney, M., and Heidmann, T. (1998). Tumor cells expressing a retroviral envelope escape immune rejection *in vivo. Proc. Natl. Acad. Sci. USA* **95,** 14920–14925.

Mayer, J., Meese, E., and Mueller-Lantzsch, N. (1997). Multiple human endogenous retrovirus (HERV-K) loci with gag open reading frames in the human genome. *Cytogenet. Cell Genet.* **78,** 1–5.

Medstrand, P., Lindeskog, M., and Blomberg, J. (1992). Expression of human endogenous retroviral sequences in peripheral blood mononuclear cells of healthy individuals. *J. Gen. Virol.* **73,** 2463–2466.

Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., Lavallie, E., Tang, X. Y., Edouard, P., Howes, S., Keith, J. C. J., and Mccoy, J. M. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403,** 785–788.

Panthier, J. J., Condamine, H., and Jacob, F. (1988). Inoculation of newborn SWR/J females with an ecotropic murine leukemia virus can produce transgenic mice. *Proc. Natl. Acad. Sci. USA* **85,** 1156–1160.

Tchenio, T., and Heidmann, T. (1991). Defective retrovirus can disperse in the human genome by intracellular transposition. *J. Virol.* **65,** 2113–2118.

Tchenio, T., and Heidmann, T. (1992). High-frequency intracellular transposition of a defective mammalian provirus detected by an in situ colorimetric assay. *J. Virol.* **66,** 1571–1578.

Tönjes, R. R., Czauderna, F., and Kurth, R. (1999). Genome-wide screening, cloning, chromosomal assignment, and expression of full-length human endogenous retrovirus type K. *J. Virol.* **73,** 9187–9195.

Urnovitz, H. B., and Murphy, W. H. (1996). Human endogenous retroviruses: Nature, occurrence, and clinical implications in human disease. *Clin. Microbiol. Rev.* **9,** 72–99.

Venables, S., Brookes, M., Fan, W., Larsson, E., Maini, R. N., and Boyd, M. T. (1995). Abundance of an endogenous retroviral envelope protein in placental trophoblast suggests a biological function. *Virology* **211,** 589–592.

Voisset, C., Bouton, O., Bedin, F., Duret, L., Mandrand, B., Mallet, F., and Paranhos-Baccala, G. (2000). Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. *AIDS Res. Hum. Retroviruses* **16,** 731–740.

Wilkinson, D. A., Freeman, J. D., Goodchild, N. L., Kelleher, C. A., and Mager, D. L. (1990). Autonomous expression of RTVL-H endogenous retrovirus-like elements in human cells. *J. Virol.* **64,** 2157–2167.

Wilkinson, D. A., Goodchild, N. L., Saxton, T. M., Wood, S., and Mager, D. L. (1993). Evidence of functional subclass of the RTLV-H family of human endogenous retrovirus-like sequences. *J. Virol.* **67,** 2981–2989.

Wilkinson, D. A., Mager, D. L., and Leong, J.-A. C. (1994). Endogenous human retroviruses. *In* "The Retroviridae" (J. A. Levy, Ed.), Vol. 3, pp. 465–535. Plenum Press, New York.