Available online at www.sciencedirect.com
ScienceDirect
Genomics 89 (2007) 1–9
GENOMICS
www.elsevier.com/locate/ygeno

Molecular genetic analyses of human endogenous retroviral elements belonging to the HERV-P family in primates, human tissues, and cancer cells ☆

Joo-Mi Yi [a,b], Kornel Schuebel [a], Heui-Soo Kim [b,*]

[a] *Cancer Biology Division, The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD 21231, USA*
[b] *Department of Biological Sciences, College of Natural Sciences, Pusan National University, Pusan 609-735, Korea*

Received 17 May 2006; accepted 22 August 2006
Available online 27 September 2006

Abstract

Human endogenous retroviral sequences family P (HERV-P) proviral sequences have been located within the human genome. Here, we identify and analyze novel putative structural genes of HERV-P in primates, human tissues, and cancer cells with an aim toward better understanding their evolutionary relationships and transcriptional potential. The expression pattern of HERV-P structural genes indicates that they are actively amplified in human tissues and widely expressed in cancer cells, suggesting a potential role in carcinogenesis. Phylogenetic analyses suggest that the HERV-P family may be divided into two distinct categories that arose during primate evolution via active gene duplication. Taken together, our data provide a better understanding of the dynamic evolutionary features and potential functional roles of the HERV-P gene family.

Keywords: HERV-P; Expression; Phylogeny; Evolution; Human tissues; Human cancer; Primates

Approximately 8% of the human genome is composed of retroviral-like structures but little is known about the origins and functions of these DNA sequences. The human endogenous retroviruses (HERVs), comprising over 200 distinct categories, are thought to have arisen from germ-line infections of ancient exogenous retroviruses approximately 10 to 50 million years ago [1,2]. During evolution, HERV sequences have been subjected to multiple amplification and transposition events resulting in a widespread distribution of complete or partial retroviral sequences throughout the human genome. It has been debated whether expansion of these HERV sequences during evolution confers beneficial biological advantages to the host, and this debate has been perhaps fueled by our limited understanding of their cellular functions [3]. Although HERV-P transcripts have been identified in human tissues [4–6] and a limited number of cell lines [7–9], little is known regarding expression patterns of HERV-P transcripts and their potential physiological roles.

HERV-P-related sequences, originally named HuRRS-P, were first identified during screens for putative retroviral primer binding sites in a human genomic library [10]. The first sequences of HuRRS-P were found to be 8.1 kb in length with a characteristic proviral structure containing 3′ and 5′ LTRs located on chromosome 7q21 and exist in a cluster of 20-40 copies throughout in the genome [10]. Recently, novel human retroviral sequences were identified during a screen of human genomic DNA using conserved probes derived from the *pol* gene of different HERV families [11]. Subsequent database analysis revealed other related sequences, which were grouped together and renamed HERV-P according to the revised nomenclature assigned to retroviral sequences [11,12].

Over the past few years, a considerable number of studies have identified various HERV families from human and primate genomes. However, little is known about the expression and

☆ Sequence data from this article have been deposited with the DDBJ/EMBL/GenBank Data Libraries under Accession Nos. AB239191–AB239248 and AB239999–AB240045.

* Corresponding author. Fax: +82 51 581 2962.
E-mail address: khs307@pusan.ac.kr (H.-S. Kim).

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.08.010

function of the HERV-P family members. Here, we identify and analyze 105 HERV-P structural genes (*gag, pol,* and *env*) and related sequences from primates, human tissues, and cancer cells.

Results and discussion

Identification of HERV-P structural genes (gag/pol/env) in primate species

We first sought to identify HERV-P structural gene elements in primates by PCR. When we analyzed samples for the presence of *pol* gene fragments, PCR products were obtained in apes (chimpanzee, bonobo, gorilla, orangutan, and gibbon), Japanese monkey (Old World monkey), night monkey (New World monkey), and lemur (prosimian), whereas the *gag* and *env* genes were not detected in any of these prosimian species (Fig. 1). HERV-W *pol* gene sequences were detected in hominoids and Old World monkeys, but not in New World monkeys, as described in a previous study [13]. From this analysis, we conclude that HERV-P sequences appear in primate lineages earlier than other HERV families and may subsequently have been amplified during primate evolution.

Expression pattern of HERV-P structural genes in human tissues and cancer cells

We also wondered whether HERV-P sequences were present in the human genome and whether these sequences were expressed in different tissues or disease states. To address this, we analyzed the expression of HERV-P genes (*gag, pol,* and *env*) in 12 human tissues and 18 cancer cell lines by reverse transcription-polymerase chain reaction (RT-PCR) and sequenced the resulting PCR products to verify their structure (Fig. 2). We identified *gag* gene expression in normal human tissues, including brain, prostate, and testis. We also detected *gag* gene expression in cancer cell lines of various origins, including brain, colon, esophagus, kidney, T cell, ovary, prostate, skin, and stomach (PFSK-1, HCT-116, TE-1, UO-31, Jurkat, OVCAR-3, PC3, LOX-IMVI, and AZ521). The *pol* gene was expressed in human brain and testis tissues, but not in prostate, heart, kidney, liver, lung, placenta, skeletal muscle, spleen, thymus, or uterus, and only in MCF7 and AZ521 of the tested cancer cells. The *env* gene was expressed specifically in human brain and several human cancer cell lines (RT4, HCT-116, Jurkat, MCF7, OVCAR-3, PC3, and AZ521). Interestingly, expression of all three HERV-P viral genes was observed in only 11 of the 18 cell lines, with differential expression in all of the sampled normal tissues, suggesting that transcriptional regulation of portions of the HERV-P genome varies in different cells.

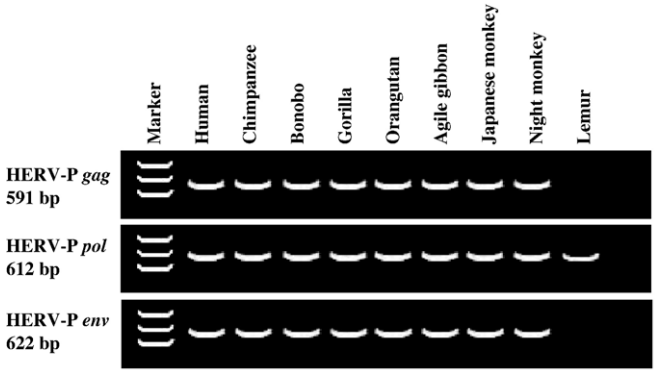

Fig. 1. Analysis of genomic DNAs for the presence of putative structural genes (*gag, pol,* and *env*) of the HERV-P family in primates. Amplification targets and product sizes are indicated on the left. DNA template species are indicated at the top.

To date, there have been several reports that HERV sequences are expressed in normal human tissues. For example, the syncytin gene, the product of the HERV-W *env* gene, is actively expressed in human placenta and testis [5] and is thought to play a role in placental cytotrophoblast fusion during development. Another HERV family member, HERV-E, was recently identified in human pancreas and thyroid tissues, suggesting a role in the physiology of these tissues, although such studies have not been undertaken [14].

Sequence analyses of the HERV-P structural genes derived from primates, human tissues, and cancer cells

The PCR and RT-PCR products derived from HERV-P sequences in various primate tissues and cell lines were cloned and sequenced. To obtain representative samples, 10 clones from each amplification reaction were randomly selected. The results of positive sequencing reactions are shown in Fig. 1. We first sequenced HERV-P structural genes, including 13 *gag* genes, 27 *pol* genes, and 18 *env* genes in primates, as shown in Table 1A. We also sequenced expressed human HERV-P sequences isolated by RT-PCR: 5 clones for *gag,* 6 clones for *pol,* and 1 clone for *env* from different human tissues and 16 clones for *gag,* 4 clones for *pol,* and 15 clones for *env* from human cancer cells (Tables 1B and 1C). Finally, we isolated, sequenced, and analyzed 58 new clones from various primate species and 47 clones from human normal tissues and cancer cell lines representing structural gene sequences belonging to the HERV-P family (74.9–100% sequence similarity to HERV-P GenBank Accession No. AC002069).

Earlier work by Kroger and Horak [10] first identified HERV-P sequences with similarity to viral *pol* genes. These results were confirmed and extended by Tristem [11]. Using sequences derived from these studies, we deduced amino acid sequences of the 37 clone HERV-P sequences isolated from primates, human tissues, and cancer cells. Intact reading frames for the 134-amino-acid HERV-P *pol* gene were found in 78% (27 of 37) of the clones (data not shown). Comparison with known *pol* genes indicated a percentage identity of HERV-P *pol* sequences ranging from 74.2 to 100%. These results are in stark contrast to the putative amino acid sequences of HERV-P *gag* and *env* gene sequences, which exhibited multiple frameshifts and termination codons caused by deleted, inserted, or mutated nucleotides.

The HERV-P structural genes showed differential expression patterns in human tissues and cancer cells. For example,