Presence and phylogenetic analysis of HERV-K LTR on human X and Y chromosomes: Evidence for recent proliferation

Heui-Soo Kim[†] and Timothy J. Crow^{*}

*POWIC, Department of Psychiatry, Warneford Hospital,
University of Oxford, Oxford, United Kingdom*

(Received 11 March, 1999, accepted 12 August 1999)

The K group of human endogenous retroviruses (HERV-K) has been suggested to have a role in disease and has recently been shown to include long terminal repeat (LTR) elements that are human specific. Here we investigated the presence of HERV-K LTRs on the human X and Y chromosomes with the use of PCR on a monochromosomal somatic cell hybrid DNA panel. We report twelve such sequences on the X chromosome and ten sequences on the Y chromosome. Phylogenetic analysis reveals that clones X2, 4, 5, 6, 7, 11, 15 from the X chromosome and clones Y4, 5, 7, 10 from the Y chromosome are closely related to the human-specific members of Medstrand and Mager's cluster 9. The sequence of clone Y7 from the Y chromosome is identical with human-specific HERV-K LTR element (AC002350) from chromosome 12q24. The findings suggest recent proliferation and transposition of HERV-K LTR elements on these chromosomes. Such events may have contributed to structural change and genetic variation in the human genome. We draw attention to evolutionarily recent changes in homologies between X and Y chromosomes as a method of further investigating such transpositions.

The HERV-K family of human endogenous retroviral sequences was originally cloned from Syrian hamster intra-cisternal A type particles (Ono et al., 1986), has homology to mouse mammary tumour virus, and includes HERV-K-related long terminal repeat (LTR) sequences that are expressed in normal placenta (Simon et al., 1994). Sequences from the HERV-K genome have been detected in humans and Old World monkeys, but not in New World monkeys by Southern blots with the HERV-K-env probe and PCR analysis (Steinhuber et al., 1995). HERV-K LTR-like elements probably entered the primate genome about 33–40 million years ago, and it is estimated that there are now approximately 25,000 copies of HERV-K LTRs in the human genome (Leib-Mosch et al., 1993).

The possibility that some of this increase has occurred recently is suggested by the presence in HERV-K sequences of a functional integrase (Kitamura et al., 1996). The ratio of synonymous to nonsynonymous substitutions indicates that evolutionary selection has recently been or is still operating (Zsiros et al., 1998). Of

particular interest is the report that some clusters (eg cluster 9 of Medstrand and Mager, 1998) of HERV-K LTRs show low (average 1.5%) divergence and that members of this cluster are specific to the human genome. HERV-K LTRs bind host cell nuclear proteins and have the potential to activate neighboring genes (Akopov et al., 1998). A possible role of HERV-K elements in disease has been considered in relation to insulin dependent diabetes mellitus (Conrad et al., 1997), seminoma (Sauter et al., 1995) and in the case of HERV-K-T47D mammary carcinoma (Seifarth et al., 1998).

The evolutionary history of retroelements in the genome is relevant to their possible role in individual differences and in pathology. The small minority of such elements that have been involved in recent retroposition is the class that is of primary interest. It has been suggested for example that such elements may play a role in the late evolutionary development of the nervous system (Crow, 1984, 1986) and that critical changes may have occurred on the sex chromosomes (Crow, 1993). Here we investigated HERV-K LTR sequences on the human X and Y chromosomes with particular reference to the class 9 elements which Medstrand and Mager (1998) have found to show low sequence divergence and to be human specific.

A human monochromosomal somatic cell hybrid DNA

^{*} Corresponding author.

[†] Present address: Division of Biological Sciences, College of Natural Sciences
Pusan National University, Pusan 609-735, Korea

panel was purchased from the Coriell Cell Repositories (Coriell Institute, USA). The X and Y chromosomal panel samples were subjected to PCR amplification. A novel HERV-K LTR family was amplified by the primer pair HS45 (5'-GTATTGTCCAAGGTTTCTCCC-3', bases 8541-8561) and DS14 (5'-GTGCTGTGCTTTTGGATATGC-3', bases 8859-8879) from the human endogenous retroviral element, HERV-K10 (GenBank, accession no. M14123). The PCR conditions followed were those of Kim et al. (1996) with an annealing temperature of 58°C. PCR products were separated on a 2% agarose gel, purified with the QIAEX II gel extraction kit (Qiagen) and cloned into the T-khs307 vector (Kim et al., 1998). The cloned DNA was isolated by the alkali lysis method using the High Pure Plasmid Isolation Kit (Boehringer Mannheim). Individual plasmid DNAs were screened for inserts by PCR. Positive samples were subjected to sequence analyses on both strands with T7 and M13 reverse primers using an automated DNA sequencer (Model 373A) and the DyeDeoxy terminator kit (Applied Biosystem). Nucleotide sequence analyses were performed using the GCG program (University of Wisconsin). Neighbor-joining phylogenetic analysis (Saitou and Nei, 1987) was performed with the MEGA program (Kumar et al., 1993). HERV-K LTR elements and SINE-R retroposons are almost homologous, but differ by a 367 bp insertion in the LTR elements (Ono et al., 1986; Kim et al., 1999). Therefore, in order to amplify HERV-K LTR elements only, PCR primers (HS45 and DS14) were chosen from the inserted sequences.

Solitary HERV-K-related long terminal repeats (LTRs) have been found to be distributed over most human chromosomes (Leib-Mösch et al., 1993). Detailed physical mapping of the HERV-K LTRs on human chromosome 19 revealed frequent proximity of the LTRs, and many have been located in the vicinity of known genes (Lebedev et al., 1995). Using PCR amplification and FISH analysis, it has been shown that gag, pol, env genes of HERV-K are found on all chromosomes except for chromosome 8, 9, 17, 18, and X (Meese et al., 1996; Mayer et al., 1997). However, a Southern analysis with probes for HERV-K-T47D related solitary LTRs using a monochromosomal somatic cell hybrid DNA panel demonstrated several hundred copies distributed over all human chromosomes (Baust, personal communication). Human-specific HERV-K LTR elements have been identified by PCR and phylogenetic analyses (Medstrand and Mager, 1998).

We identified 22 HERV-K LTR elements (12 from the X and 10 from the Y) from the human sex chromosomes, and subjected them to phylogenetic analysis with the human-specific HERV-K LTRs found by Medstrand and Mager (1998) (Fig. 1). Clones X2, 4, 5, 6, 7, 11, 15 from the X chromosome and clones Y4, 5, 7, 10 from the Y chromosome are closely related to the human-specific members of Medstrand and Mager's cluster 9. The sequence of clone

Y7 from Y chromosome is identical with human-specific HERV-K LTR element (AC002350) from chromosome 12q24, and this sequence differs from AC002400, Y4 and Y5, L47334 and AC002508, by one, two or three bp respectively. The finding suggest that there has been recent proliferation and transposition of HERV-K LTR elements. Such events may have contributed to structural change and genetic variation in the human genome.

The nucleotide sequences of the HERV-K LTR isolated from the sex chromosomes have at least 92% sequence identity with the HERV-K10 LTR. Although HERV-K elements have been previously detected on the Y chromosome by Southern analysis (Mayer et al., 1997), this appears to be the first report of such sequences on the X chromosome. Clones Y4 and Y5 have 99.3% sequence identity with clone Y7. An average of 1.5% divergence reveals these sequences as belonging to the human-specific cluster (Medstrand and Mager, 1998). Similarly AC002400 has over 98.6% sequence identity with X5, 6 and 11 as well as with L47334, AC002508, U73641, Z80898 and U47924 within Medstrand and Mager's cluster 9, and U73641 differs from the sequence of HERV-K10 LTR by only one base pair. This group (X2, 4, 5, 6, 7, 11, 15 and Y4, 5, 7, 10) represents an expansion on the X and Y chromosomes that has occurred since the separation of the chimpanzee lineage.

The relationships between HERV-K LTR elements on the human X and Y chromosomes is of interest. Although there appear to be relatively separate sub-clusters on the X (X5, 6, 11 and X4, 7, 15) and on the Y (Y4, 7 and 10), there is also overlap (X2 and Y5) (Fig. 1). One possibility is that some part of the mixed X and Y group relates to regions of homology that have been established late in evolution such as the 4 Mb block in Yp that transposed from Xq21.3 after the separation of the chimpanzee and hominid lineages (Sargent et al., 1996; Mumm et al., 1997) and was subsequently split by a paracentric inversion on the Y chromosome. The transposition has been dated at 3 to 4 million years, and the paracentric inversion took place after this but before the radiation of extant human populations (Schwartz et al., 1998). These two blocks in Yp are thus specific to the human genome. If HERV-K elements are present within the Xq21.3 region or either of the homologous blocks in Yp sequence analysis will cast further light on their evolutionary history. Elements that were already present on the X chromosome at the time of the transposition will be represented on the Y chromosome. Discrepancies in the sequences on the X and the Y must have arisen in the past 3 to 4 million years, and there is the possibility of detecting new insertions or deletions. The existence of X-Y homologous regions together with their evolutionary chronology (Lambson et al., 1992) therefore provides an approach to the timing of retrotransposition events.

Fig. 1. Phylogenetic tree obtained by the neighbor-joining method for HERV-K LTR elements including human-specific HERV-K LTRs as described by Medstrand and Mager (1998). The accession numbers (AC002350, AC002400, AC002508, L47334, U47924, U73641, Z80898) of human-specific HERV-K LTRs were obtained from the GenBank database with the aid of the BLAST network server (Altschul et al., 1997). The nucleotide sequence of HERV-K10 LTR was taken from GenBank accession no. M14123. Branch lengths are proportional to the distance between the taxa. Statistical robustness was tested by bootstrap analysis - numbers to the left of the nodes of the tree branches represent the number of times out of 100 trees. The nucleotide sequence data reported in this paper will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB022739 - AB022760.

This work was supported by grants from the UK Medical Research Council and the Stanley Foundation.

REFERENCES

- Akopov, S. B., Nikolaev, L. G., Khil, P. P., Lebedev, Y. B., and Sverdlov, E. D. (1998). Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. *FEBS Lett.* **421**, 229–233.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
- Conrad, B., Weissmahr, R. N., Boni, J., Arcari, R., Schupbach, J., and Mach, B. (1997). A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* **90**, 303–313.
- Crow, T. J. (1984). A re-evaluation of the viral hypothesis: is psychosis the result of retroviral integration at a site close to the cerebral dominance gene? *Br. J. Psychiatry* **145**, 243–253.
- Crow, T. J. (1986). Left brain, retrotransposons, and schizophrenia. *Br. Med. J.* **293**, 3–4.
- Crow, T. J. (1993). Sexual selection, Machiavellian intelligence and the origins of psychosis. *Lancet* **342**, 594–598.
- Kim, H.-S., Chen, Y., and Lonai, P. (1998). Complex regulation of multiple cytohesin like genes in murine tissues and cells. *FEBS Lett.* **433**, 312–316.
- Kim, H.-S., Hirai, H., and Takenaka, O. (1996). Molecular features of the TSPY gene of gibbons and Old World monkeys. *Chrom. Res.* **4**, 500–506.
- Kim, H.-S., Wadekar, R. V., Takenaka, O., Winstanley, C., Mitsunaga, F., Kageyama, T., Hyun, B.-H., and Crow, T. J. (1999). SINE-R.C2 (a *Homo sapiens* specific retroposon) is homologous to cDNA from post-mortem brain in schizophrenia and to two loci in the Xq21.3/Yp block linked to handedness and psychosis. *Am. J. Med. Genet. (Neuropsychiat. Genet.)* **88**, 560–566.
- Kitamura, Y., Ayakuwa, T., Ishikawa, T., Kanda, T., and Yoshiike, K. (1996). Human endogenous retrovirus K10 encodes a functional integrase. *J. Virol.* **70**, 3302–3306.
- Kumar, S., Tamura, K., and Nei, M. (1993). MEGA: Molecular evolutionary genetics analysis, version 1.01. The Pennsylvania State University, University Park, PA 16802.
- Lambson, B., Affara, N. A., Mitchell, M., and Ferguson-Smith, M. A. (1992). Evolution of DNA sequence homologies between the sex chromosomes in primate species. *Genomics* **14**, 1032–1040.
- Lebedev, Y. B., Volik, S. V., Obradovic, D., Ermolaeva, O. D., Ashworth, L. K., Lennon, G. G., and Sverdlov, E. D. (1995). Physical mapping of sequences homologous to an endogenous retrovirus LTR on human chromosome 19. *Mol. Gen. Genet.* **247**, 742–748.
- Leib-Mösch, C., Haltmeier, M., Werner, T., Geigl, E.-M., Brack-Werner, R., Francke, U., Erfle, V., and Hehlmann, R. (1993). Genomic distribution and transcription of solitary HERV-K LTRs. *Genomics* **18**, 261–269.
- Mayer, J., Meese, E., and Mueller-Lantzsch, N. (1997). Chromosomal assignment of human endogenous retrovirus K (HERV-K) *env* open reading frames. *Cytogenet. Cell Genet.* **79**, 157–161.
- Medstand, P., and Mager, D. L. (1998). Human-specific integrations of the HERV-K endogenous retrovirus family. *J. Virol.* **72**, 9782–9787.
- Meese, E., Gottert, E., Zang, K. D., Sauter, M., Schommer, S., and Mueller-Lantzsch N (1996) Human endogenous retroviral element K10 (HERV-K10): chromosomal localization by somatic hybrid mapping and fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **72**, 40–42.
- Mumm, S., Molini, B., Terrell, J., Srivastava, A., and Schlessinger, D. (1997). Evolutionary features of the 4-Mb Xq21.3 XY homology region revealed by a map at 60-kb resolution. *Genome Res.* **7**, 307–314.
- Ono, M., Yasunaga, T., Miyata, T., and Ushikubo, H. (1986). Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. *J. Virol.* **60**, 589–598.
- Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- Sargent, C. A., Briggs, H., Chalmers, L. J., Lambson, B., Walker, E., and Affara, N., (1996). The sequence organization of Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. *Genomics* **32**, 200–209.
- Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, I., Buck, M., Best, B., Neumann-Haefelin, D., and Mueller-Lantzsch, N. (1995). Human endogenous retrovirus K10: expression of gag protein and detection of antibodies in patient with seminomas. *J. Virol.* **69**, 414–421.
- Schwartz, A., Chan, D. C., Brown, L. G., Alagappan, R., Pettay, D., Distech, C., McGillivray, B., De la Chapelle, A., and Page, D. C. (1998). Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**, 1–11.
- Seifarth, W., Baust, C., Murr, A., Skladny, H., Krieg-Schneider, F., Blusch, J., Werner, T., Hehlmann, R., and Leib-Mösch C. (1998): Proviral structure, chromosomal location, and expression of HERV-K-T47D, a novel human endogenous retrovirus derived from T47D particles. *J. Virol.* **72**, 8384–8391.
- Simon, M., Haltmeier, M., Papakonstantinou, G., Werner, G., Hehlmann, R., and Leib-Mösch, C. (1994). Transcription of HERV-K-related LTRs in human placenta and leukemic cells. *Leukemia* **8**, S12–17.
- Steinhuber, S., Brack, M., Husmann, G., Schwelberger, H., Dierich, M. P., and Vogetseder, W. (1995). Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. *Hum. Genet.* **96**, 188–192.
- Zsiros, J., Jebbink, M. F., Lukashov, V. V., Voute, P. A., and Berkhout, B. (1998). Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses. *J. Gen. Virol.* **79**, 61–70.