Presence and Phylogenetic Relationships of a Hominoid-Specific Retroposon Family on the Human Y Chromosome

Authors: Kim, Heui-Soo, and Crow, Timothy J.

Source: Zoological Science, 16(6) : 963-970

Published By: Zoological Society of Japan

URL: https://doi.org/10.2108/zsj.16.963

ZOOLOGICAL SCIENCE **16**: 963–970 (1999)

Presence and Phylogenetic Relationships of a Hominoid-Specific Retroposon Family on the Human Y Chromosome

Heui-Soo Kim and Timothy J. Crow*

POWIC, Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford, OX3 7JX, United Kingdom

ABSTRACT—The retroposon SINE-R.C2 was first identified as a human-specific insertion in the complement C2 gene and the insertion of a closely related element has been implicated as a cause of Fukuyama type muscular dystrophy. In a previous study, we found SINE-R-type retroposons, derived from the endogenous retrovirus HERV-K family, have been found to be hominoid specific. In this report, we identified twelve SINE-R-type retroposons from the human Y chromosome by a PCR approach. Compared with other human retroposons (SINE-R.C2, 11, 14, 19, and HS307/HS408), the Y-chromosomal retroposon showed a high degree of sequence homology (86–96%). Phylogenetic analysis using the neighbor-joining method revealed that SINE-R-type retroposons on the Y chromosome were inter-related with hominoid primates, suggesting that SINE-R elements on the Y chromosome have evolved progressively in the hominoid radiation. Two elements SINE-R.Y-2 and Y-7 are more closely related to the human-specific retroposon SINE-R.C2 than other Y-chromosomal retroposons. Such elements thus are likely to be of great potential relevance to recent events in hominoid evolution and to structural change or genetic variation connected to various disease.

INTRODUCTION

It has often been suggested that endogenous retroviruses and retroposons have played a role in influencing the functional organization of the human genome (Baltimore, 1985; Kazazian and Moran, 1998; Sverdlov, 1998). Approximately 1% of the human genome is comprised of such sequences, and most of them have been integrated into the primate lineage thirty-three or more million years ago (Haltmeier *et al*., 1995; Mager and Freeman, 1995; Steinhuber *et al*., 1995). The vast majority of these elements were incorporated as proviral DNA which then rapidly became replication-defective by random accumulation of mutations to be vertically transmitted as junk DNA. Therefore, particular interest attaches to families of elements that have changed recently in human and hominoid evolution. By PCR and phylogenetic analyses, Medstrand and Mager (1998) found that some solitary long terminal repeat (LTR) elements of the human endogenous retrovirus K family (HERV-K) were human specific. These elements may still be actively transposing in the human genome. For example, a HERV-K LTR element has been inserted in the antisense orientation between the gag and pol gene of the human endogenous retrovirus S71 (Leib-Mösch *et al*., 1993). Another solitary element has also been found at the HLA-DQB1 (Seidl *et al*., 1999). It has been estimated that there are 25,000 copies of such elements in the human genome (Leib-Mösch *et al*., 1993), and they are potentially capable of affecting the expression of closely located genes (Akopov *et al*., 1998). Simon *et al*. (1994) found that HERV-K-related LTR elements are expressed in human placenta.

Amongst the derivatives of the HERV-K10 retrovirus family is a family of Short Interspersed Nuclear Elements (SINEs) termed SINE-R for retrovirus-derived (Ono *et al*., 1987). SINE-R11, 14, 19 were isolated by colony blot hybridization with a HERV-K10 probe, and derive from the 3' long terminal repeat and small upstream flanking regions. These elements entered in the genome of hominoid primates after the split of Old World monkeys in the Oligocene period (Kim *et al*., 1999a, b). The SINE-R.C2 element was discovered in relation to a variable number of tandem repeat (VNTR) sequence within the third intron of the gene for C2, the second component of complement, located in class III of the major histocompatibility complex on the short arm of human chromosome 6 (Carroll *et al*., 1984; Zhu *et al*., 1992). By Southern blot analysis, SINE-R.C2 is confined to Homo sapiens (Zhu *et al*., 1994). Within the block in Xq21.3 with homology to Yp11.2 that has been tentatively linked to shizophrenia and schizo-affective disorder (Laval *et al*., 1998) we previously reported two retroposons (HS307 and HS408) with 92 to 96% homology to SINE-R.C2 (Kim *et al*., 1999b). An element almost identical in sequence

* Corresponding author: Tel. +44-1865-226474;
FAX. +44-1865-244990.
E-mail: tim.crow@psychiatry.oxford.ac.uk

to SINE-R.C2 has recently been reported as the cause of Fukuyama-type muscular dystrophy (Kobayashi *et al.*, 1998), and this appears to be the first association of this class of elements with disease. Influences of SINE-R elements on gene expression have been described, for example, a SINE element that acquired a role in signal transduction in the course of mammalian evolution (Shimamura *et al.*, 1998). Here we report the presence of SINE-R-type retroposons on the human Y chromosome and their phylogenetic analysis with those of other humans and hominoids.

MATERIALS AND METHODS

Sample DNA and PCR

A human monochromosomal somatic cell hybrid DNA panel was obtained from the MRC Human Genome Mapping Project Resource Centre (HGMP-RC, Cambridge, UK, Kelsall *et al.*, 1995). The Y chromosomal DNA sample was subjected to PCR amplification. SINE-R-type retroposons were amplified by the primer pair JS902 (5'-GAGAATGGGCCATGATGAC-3', bases 4–22) and JS903 (5'-GATCATTCTTGGATGTTTCT-3', bases 466–485) from schizophrenic brain S11 cDNA (GenBank, accession no. AA772777) which has a high level of sequence homology with SINE-R.C2, 11, 14, 19, and HS307/HS408 elements (Ono *et al.*, 1987; Kim *et al.*, 1999b). The PCR conditions followed were those of Kim *et al.* (1996) with an annealing temperature of 56°C.

Cloning of PCR product

PCR products were separated on a 1.8% agarose gel, purified with the QIAEX II gel extraction kit (Qiagen) and cloned into the T-khs307 vector (Kim *et al.*, 1998). Plasmid DNA was isolated by the alkali lysis method using a High Pure plasmid isolation kit (Boehringer Mannheim).

DNA sequencing and data analysis

Individual plasmid DNAs were screened for inserts by PCR. Positive samples were subjected to sequence analysis on both strands with T7 and M13 reverse primers using an automated DNA sequencer (Model 373A) and the DyeDeoxy terminator kit (Applied Biosystem). SINE-R.C2 (L09706), SINE-R11 (X07417), SINE-R14 (X07418), SINE-R19 (X07419), FMR-1 (L29074), HS307 (AB016781), HS408 (AB016782), SCHIZO (AA772777), Chi-M (AB018364), Chi-F (AB018365), Gor-M (AB018366), Gor-F (AB018367), Ora-M (AB018368), Ora-F (AB018369), Gib-M (AB018370), and Gib-F (AB018371) sequences were taken from the GenBank database. Sequence analyses were done with the aid of GAP, PILEUP, and PRETTY from the GCG program (University of Oxford). Neighbor-joining phylogenetic analysis (Saitou and Nei, 1987) was performed with the CLUSTAL W program (Thompson *et al.*, 1994) Sequence homology was searched for in the BLAST network server (Altschul *et al.*, 1997).

RESULTS AND DISCUSSION

The nucleotide sequences of the human endogenous retrovirus, HERV-K10, have been found to be present in hominoids and Old World monkeys by PCR analysis using genomic DNA (Zhu *et al.*, 1994). SINE-R-type retroposons, that are clearly derived from the long terminal repeat (LTR) of the HERV-K family (Ono *et al.*, 1987; Zhu *et al.*, 1992), have been found to be hominoid-specific by PCR analysis (Kim *et al.*, 1999b). Comparing nucleotide sequences between SINE-R-type retroposons and HERV-K LTR elements, we showed that the 367 bp of SINE-R-type retroposons are deleted from the HERV-K LTR sequences (Ono *et al.*, 1987; Kim *et al.*, 1999b). By BLAST search, cDNA prepared from the brain of a patient with schizophrenia (GenBank, accession no. AA772777) was found to have sequences containing a high degree of homology with the 3' LTR of HERV-K10 (Ono *et al.*, 1986), SINE-R11, 14, 19 (Ono *et al.*, 1987), SINE-R.C2 (Zhu *et al.*, 1994), and HS307/HS408 (Kim *et al.*, 1999b). Therefore, PCR primers, JS902 and JS903, were chosen from the brain cDNA of a patient with schizophrenia. Using genomic DNA from the human Y chromosome, we identified twelve SINE-R-type retroposons by PCR amplification in the present study although we are aware that there are more SINE-R-type retroposons on the Y chromosome. Ono *et al.* (1987) have estimated that SINE-R elements were present at 4000 to 5000 copies per haploid human genome. So far, limited numbers of SINE-R-type retroposons (SINE-R11, 14, 19, C2, and HS307/HS408) have been reported in the human genome (Ono *et al.*, 1987; Zhu *et al.*, 1994; Kim *et al.*, 1999b). As far as we know, this is the first time that the nucleotide sequences of Y-chromosomal retroposons have been demonstrated. The sequences are presented in alignment with those of the other human retroposon family in Fig. 1. The sequence data from the human Y chromosome was demonstrated as derived from SINE-R-type retroposons by comparison with other human retroposons. The hominoid specificity of the PCR products was also demonstrated by the absence of amplified rodent DNA (data not shown). As shown in Table 1, Y-chromosomal retroposons have a high degree of sequence homology with other retroposons. Sequence similarity ranged from 99.5% between SINE-R.Y-7 and SINE-R.Y-2, also between SINE-R.Y-15 and Y-17, to 85.6% between SINE-R.Y-4 and SINE-R.Y-15. The human-specific retroposon SINE-R.C2 shared 96% sequence homology with both SINE-R.Y-2 and Y-7.

To understand the evolutionary relationships among these families with hominoid retroposons, a phylogenetic tree was constructed by the neighbor-joining method. As shown in Fig. 2, the sequence of the chimpanzee male (Chi-M) is closely related to the sequences from SINE-R.Y-8 and SINE-R.Y-4, while SINE-R.Y-15 and Y-17 cluster with the sequence from the gibbon male (Gib-M) and female (Gib-F), the schizophrenic brain cDNA (SCHIZO), and the orangutan female (Ora-F). Retroposons Y-3, 11, 13, 16, 18, 19 are closely related to SINE-R14. SINE-R.Y-7 and SINE-R.Y-2 are more closely related to each other than either of them is to SINE-R11 and they cluster with SINE-R.C2. The human-specific retroposon SINE-R.C2 is closely related to the sequence of gorilla male (Gor-M) and these sequences cluster with the HS307 retroposon that we previously identified in the Xq21.3 region of homology with the Y chromosome (Kim *et al.*, 1999b). This group with SINE-R.C2 is likely to be of great relevance to recent events in hominoid evolution. Thus the sequences of SINE-R elements on the Y chromosome are intermingled with those of similar elements in hominoid primates, suggesting that several lineages of retroposons have been evolving independently during hominoid evolution.

The non-recombining portion of the human Y chromo-

```
CONSENSUS  ATGGCGGTTT TGT-GAATAG AAAAGGGGGA AATGTGGGGA AAAGATAGAG
           1                                                   50
SINE-RC2   ---------- ---G------ ---G------ --G------- ------T---
SINE-R11   ---------- ---G------ ---------- --G------- ------T---
SINE-R14   ---------- ---C------ ---G------ ---------- G----GG---
SINE-R19   ---------- ---C------ ------T--- ---------- -------C--
SCHIZO     ---------- ---C---A-- ---------- ---------- -----A----
FMR-1      -----A-C-- ---G--G--- ---G-C---- --G------- -GG---T---
HS307      ------T--- ---G----G- ---G------ --G------- ------T---
HS408      -----A---- ---C--G--- -G-G------ ---------- GGG---G---
Y-2        ---------- ---G------ ---------- --G------- ------T---
Y-3        ---------- GCCGA-..-A G--------- ---------- -----A----
Y-4        ---------- ---G------ ---G------ --G------- ------G---
Y-7        ---------- ---G------ ---------- --G------- ------T---
Y-8        ---------- ---G------ ---G------ --G------- ------G---
Y-11       G--------- ---C---A-- ---------- ---------- -----A----
Y-13       ---------- ---C------ ---------- ---------- -----A----
Y-15       ---------- --CC---A-- ---------- ---------- -----A----
Y-16       ---------- ---C------ ---------- ---------- -----A----
Y-17       ---------- --CC---A-- ---------- ---------- -----A----
Y-18       ---------- ---C------ ---------- ---------- -----A----
Y-19       ------A--- ---GA----- ---------- -------A-- -----A----

CONSENSUS  A-ATCAGATT GTT-CTGTGT CTGTGTAGAA AGAAGTAGAC ATGGGAGACT
           51                                                  100
SINE-RC2   -A---G---G ---G-C---- ---------- ---G------ C---------
SINE-R11   -A---G---G ---G------ ---------- ---------- ----------
SINE-R14   -G---G-T-- ---A------ ------G--- G--G------ --TA------
SINE-R19   -A-------A ---G------ ---------- ---------- --A--G--T-
SCHIZO     -G-------- ---A---C-- ---------- ---------- --A-------
FMR-1      -AG------A ---G-C---- ------G--- G--------- ----------
HS307      -A-A-G---G ---G-C---- ---------- ---------- -------T--
HS408      -GG------- ---G------ ---------G G--------- --A---T---
Y-2        -A---G---G ---G------ -----A---- ---------- ----------
Y-3        -G-------- T--A------ -----C---- ---------- --A------A
Y-4        -A-------- ---G------ ---------- -----C---- ---------.
Y-7        -A---G---G ---G------ -----A---- ---------- ----------
Y-8        -A-------- ---G------ ---------- -----C---- ---------.
Y-11       -A-------- ---A------ ---------- ---------- ---A------
Y-13       -G-------- ---A------ -------A-- ---------- --A--G----
Y-15       -G-------- T--A------ -----C---- ---------- --A------A
Y-16       -G-------- ---A------ -------A-- ---------- --A--G----
Y-17       -G-------- T--A------ -----C---- ---------- --A------A
Y-18       -G-------- ---A------ -------A-- ---------- --A--G----
Y-19       -G-------- ---A------ --------C- ---------- ----------

CONSENSUS  -CACTTTGTT GTTCTGTACT AAGAAAAATT CTTCTGCCTT GGGATGCTGT
           101                                                 150
SINE-RC2   TTT-A--.-- ---------- ---------- ---------- -----C----
SINE-R11   TTT-A--.-- ---------- ---------- -----T---- -----C----
SINE-R14   C--T----.. .--------- -G-------- ---------- ----------
SINE-R19   C--T----.. .--------- ---------- ---------- ----------
SCHIZO     C--T----.. .--------- ---------- ---------- ----------
FMR-1      TTT-A--.-- ------C--- GG--G----- -C-------- -----C----
HS307      TTT-A--.-- ---------- ---------- --------G- -----C----
HS408      C--T----.. .--------- ---------- ---------- ----------
Y-2        TTT-A--.-- ---------- ---------- ---------- -----C----
Y-3        C-------.. .--------- ---------- ---------G ----------
Y-4        .TT-A--.-- --C------C -G--GGGG-- --G------- ----------
Y-7        TTT-A--.-- ---------- ---------- ---------- -----C----
Y-8        .TT-A--.-- ---------- GG--GGGG-- --G------- ----------
Y-11       C--T----.. .--------- ---------- ---------- ----------
Y-13       T--T------ C--------- ---------- ---------- ----------
Y-15       C-------.. .--------- ---------- ---------G ----------
Y-16       T--T------ C--------- ---------- ---------- ----------
Y-17       C-------.. .--------- ---------- ---------G ----------
Y-18       T--T------ C--------- ---------- ---------- ----------
Y-19       C--T----.. .--------- ---------- ---------- -------CA-
```

```
CONSENSUS  TAATCTATGA CCTTACCCCC AACCCTGTGC TCTCTGAAAC ATGTGCTGTG
           151                                                 200
SINE-RC2   -G---GG--- ---------- ---------- ---------- ---------A
SINE-R11   -G-------- ---------- ---------- ---------- ----------
SINE-R14   --------A- ---------- -----C---- ---------- ----------
SINE-R19   -G-------- ---------- -----G---- -G-------- ----------
SCHIZO     --------A- ---------- -----CC--- ---------- ----------
FMR-1      -G----G--- ----G----- ---------- ---------- ----------
HS307      GG----G--- ---------- ---------- ---------- ----------
HS408      --------A- ---------- ---------- ---------- -A--------
Y-2        -G-------- ---------- ---------- ---------- ------A---
Y-3        --------A- --A-G----- -----C---- ---------- ----------
Y-4        -G----G--- ----G----- -G-------- ------GGG- ---------A
Y-7        -G-------- ---------- ---------- ---------- ----------
Y-8        -G----G--- ----G----- -G-------- ------GGG- ----------
Y-11       --------A- ---------- ---------- ---------- ----------
Y-13       ---------- ---------- -----C---- ---------- -G--------
Y-15       --------A- --A------- -----C---- ---------- ----------
Y-16       ---------- ---------- -----C---- ---------- -G--------
Y-17       --------A- --A------- -----C---- ---------- ----------
Y-18       ---------- ---------- -----C---- ---------- -G--------
Y-19       --------A- ---------- -----CC--T ---------- ----------

CONSENSUS  TCCACTCAGG GTTAAATGGA TTAAGGGCGG TGCAAGATGT GCTTTGTTAA
           201                                                 250
SINE-RC2   ---------- ---G------ -----A---- ---------- ----------
SINE-R11   ---------- ---------- ---------- ---------- ----------
SINE-R14   ---------- ------AT-- ---------- ---------- ----------
SINE-R19   ---------- ---------- ---------- ---------- ----------
SCHIZO     --A------- ---------- ---------- ---------- ----------
FMR-1      ---------- ---------- ---------- ---------- ----------
HS307      ---------- ---------- ---------- ---------- ----------
HS408      ---------- ---------- ---------- ---------- ----------
Y-2        ---------- ---------- ---------- -----.---- ----------
Y-3        --A------- ---------- ---------- ---------- ----------
Y-4        ---------- ---------- -------T-- ---------- ----------
Y-7        ---------- ---------- ---------- -----.---- ----------
Y-8        ---------- ---------- -------T-- ---------- ----------
Y-11       ---------- ---------- -------T-- ---------- C---------
Y-13       ---------- ---------- ---------- -------C-- ----------
Y-15       --A------- ---------- ---------- ---------- ----------
Y-16       ---------- ---------- ---------- ---------- ---C------
Y-17       --A------- ---------- ---------- ---------- ----------
Y-18       ---------- ---G------ ---------- ---------- ----------
Y-19       ---------- ---------- ---------- ---------- ----------

CONSENSUS  ACAGATGCTT GAAGGCAGCA TGCTCGTTAA GAGTCATCAC CACTCCCTAA
           251                                                 300
SINE-RC2   ---------- ---------- -----C---- ---------- ----------
SINE-R11   ---------- ---------- ---------- ---------- ----------
SINE-R14   ---------- ---------- ---------- ---------- ----------
SINE-R19   ---------- ---------- ---A-T---- -----C---- ----------
SCHIZO     ---------- ----A----- -----C---- ---------- ----------
FMR-1      ---------- ---------- ---------- ---------- --A-------
HS307      ---------- ---------- -----C---- ---------- ----------
HS408      ---------- ---------- -----A---- ---------- ----------
Y-2        ---------- ---------- -----A---- ---------- ----------
Y-3        ---------- ---------- -----T---- ---------- ----------
Y-4        --------C- ---------- ---------- ---------- ----------
Y-7        ---------- ---------- -----A---- --------G- ----------
Y-8        ---------- ---------- ---------- ---------- ----------
Y-11       ---------- ---------- -------G-- ---------- ----------
Y-13       ---------- -----TG--- ---------- ---------- ----------
Y-15       ---------- A--------- ---------- ---------- ----T-----
Y-16       ---------- -----TG--- ---------- ---------- ----------
Y-17       ---------- A--------- ---------- ---------- ----------
Y-18       ---------- -----TG--- ---------- ---------- ----T-----
Y-19       ---------- ---------- -----T---- ---------- ----------
```

```
CONSENSUS  TCTCAAGTAC CCAGGGACAC AAACACTGCG GAAGGCCGCA GGGTCCTCTG
           301                                                 350
SINE-RC2   ---------- ---------- ---------- ---------- ----------
SINE-R11   ---------- ---------- ---------- ---------- ----------
SINE-R14   ------A--- ---------- --------T- ---------- ----------
SINE-R19   ---------- ---------- ---------- ---------C ----------
SCHIZO     ---------- ---------T -----GG--T ---------- A--A------
FMR-1      ---------A T--------- ---------T ---------- ----------
HS307      -----T---- ---------- ---------- ---------- ----------
HS408      ---------- ---------- ---------C ------T--- ----------
Y-2        ---------- ---------- --------T- ---------- ----------
Y-3        ---------- ---------- ----C----. -------A-- ----------
Y-4        ---------- ---------- ---------A ------T--- ----------
Y-7        ---------- ---------- --------T- ---------- ----------
Y-8        ---------- ---------- ---------A ------T--- ----------
Y-11       ---------- ---------- ---T------ ---------- ----------
Y-13       ---------- -----TG--- ---------- ----C----- ----------
Y-15       ---------- ---------- ------A--C ---------- --AA------
Y-16       ---------- -----TG--- ---------- ----C----- ----------
Y-17       ---------- ---------- ------A--C ---------- --AA------
Y-18       ---------- -----TG--- ---------- ----C----- ----------
Y-19       ---------- ---------- ----C----. -------A-- ----------

CONSENSUS  CCTAGGAAAA CCAGAGACCT TTGTTCACTT GTTTATCTGC TGACCTTCCC
           351                                                 400
SINE-RC2   ---------- ---------- ---------- ---------- ----------
SINE-R11   ---------- ---------- ---------- ---------- ----------
SINE-R14   ---------- ---------- --------A- ---------- ----------
SINE-R19   ---------- ---------- G--------- ---------- ----T-----
SCHIZO     ---------- ---------- -------TG- ---------- -------TT-
FMR-1      ---------- ---------- ---------- ---------- ----------
HS307      ---------- -------T-- ---------- ---------- -----C----
HS408      ---------- ---------- ---------- ---------- ---T------
Y-2        ---------- ---------- ---------- ---------- ---T------
Y-3        ---------- --G------C --------A- A--------- ----------
Y-4        ---------- A-CAGA-A-- ---------- ---------- ----------
Y-7        ---------- ---------- ---------- ---------- ---T------
Y-8        ---------- ---------- ---------- ---------- ----------
Y-11       ---------- ---------C --------A- ----G----- ----------
Y-13       ---------- ---------- -------TA- ---------- --------T-
Y-15       ---------- G--------- --------G- ---------- --------TT
Y-16       ---------- ---------- -------TA- -----C---- --------T-
Y-17       ---------- G--------- --------G- ---------- --------T-
Y-18       ---------- ---------- -------TA- ---------- -------TT-
Y-19       ---------- --G------C --------A- A--------- ----------

CONSENSUS  TCCACTATTG TCCTATGACC CTGCCAAATC CCCCTCTCTG --
           401                                       442
SINE-RC2   ---------- ----G----- ---------- -------G-- ..
SINE-R11   ---------- ---------- ---------- -------G-- ..
SINE-R14   ---------- ---------- ---------- ---------- ..
SINE-R19   -----G---- ---------A --A------- -------GC- ..
SCHIZO     ---------A ---------- ------C--- --------C- ..
FMR-1      ---------- ---C------ ---------- -------G-- ..
HS307      ----T----- ---------- ---------- -------GC- ..
HS408      ---------- ---------- ---------- -------GC- ..
Y-2        ---------- ---C------ ---------- ---G---G-- ..
Y-3        ---------- ---------- ---------- ---------- ..
Y-4        ------T-C- ---------- ---------- ------CTGT GC
Y-7        ---------- ---C------ ---------- ---G---G-- ..
Y-8        ---------- ---------- ---------- -------G-- ..
Y-11       ---------A ---------- ---T--C--T ---------- ..
Y-13       ---------A ---------- ------C--- ---------- ..
Y-15       ---------A ---------- ------C--- ---------- ..
Y-16       ---------- ---------- ---------- ---------- ..
Y-17       ---------A ---------- ------C--- ---------- ..
Y-18       ---------- ---------- ---------- ---------- ..
Y-19       ---------- ---------- ---------- ---------- ..
```

Fig. 1. Nucleotide sequence alignments of the retroposon family. Consensus sequences are shown on the top row. Dashes indicate no change to the consensus sequence and dots indicate gaps. The nucleotide sequence data reported in this paper for Y-chromosomal retroposons will appear in the DDBJ/EMBL/GenBank nucleotide sequence databases with the accession numbers AB025287–AB025298.

Table 1. Percentage similarity of nucleotide sequences in the retroposon family

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. SINE-RC2	–																			
2. SINE-R11	97.3	–																		
3. SINE-R14	92.2	93.6	–																	
4. SINE-R19	92.4	94.5	91.1	–																
5. SCHIZO	90.2	92.0	92.0	91.1	–															
6. FMR-1	93.2	94.1	90.2	89.9	87.6	–														
7. HS307	95.9	96.4	90.8	92.0	89.7	92.3	–													
8. HS408	91.8	93.4	93.1	92.0	91.5	92.0	91.3	–												
9. Y-2	96.1	98.2	92.7	93.3	90.8	93.2	95.2	92.9	–											
10. Y-3	89.6	91.7	91.7	90.3	92.4	88.0	89.2	90.8	90.8	–										
11. Y-4	90.4	91.3	87.9	87.6	86.0	89.0	89.5	88.3	89.9	86.2	–									
12. Y-7	96.1	98.2	92.7	93.3	90.8	93.2	95.2	92.9	99.5	90.8	89.9	–								
13. Y-8	93.4	94.7	91.1	90.8	89.0	92.9	92.4	91.5	93.3	89.2	96.1	93.3	–							
14. Y-11	91.5	93.8	93.4	91.5	93.4	89.0	90.6	91.5	92.4	92.6	87.9	92.4	91.1	–						
15. Y-13	90.2	92.5	92.9	92.0	94.1	87.7	89.3	91.5	91.1	91.9	87.2	91.1	90.4	93.8	–					
16. Y-15	89.2	91.5	91.5	89.9	94.5	87.0	88.3	90.6	90.4	94.7	85.6	90.4	88.6	92.9	93.1	–				
17. Y-16	90.4	92.7	93.1	92.2	93.4	87.9	89.5	91.8	91.3	92.2	87.4	91.3	90.6	93.1	98.9	92.4	–			
18. Y-17	89.7	92.0	92.0	90.4	95.0	87.4	88.8	91.1	90.8	95.2	86.0	90.8	89.0	93.4	93.6	99.5	92.9	–		
19. Y-18	90.7	92.5	93.0	92.0	93.6	87.7	89.3	91.5	91.1	91.9	87.2	91.1	90.4	92.9	98.6	92.7	98.9	92.7	–	
20. Y-19	91.3	93.3	92.9	91.3	92.9	88.8	90.6	91.5	92.2	95.2	87.4	92.2	90.4	93.6	92.9	91.7	93.1	92.2	92.9	–

some includes relatively few genes and has the capacity to accumulate repetitive and other junk DNA sequences unconstrained by the functional considerations that apply elsewhere. It shows considerably greater variation between hominoid species (Glaser *et al*., 1997). However, particular interest attaches to regions of homology between the X and Y chromosomes. Such homologies have generally been created by X to Y transpositions and a number of such events have been tracked in the course of mammalian and primate evolution (Lambson *et al*., 1992; Glaser *et al*., 1997). Genes within these regions are thus exposed to new evolutionary pressures, and their X homologues are subject to differential patterns of inactivation (Jegalian and Page, 1998). Some evolutionary changes are known to have been influenced by retroposons, eg the LINE-LINE recombination associated with the paracentric inversion on the Y chromosome (Schwartz *et al*., 1997) and they may be the most recent chromosomal rearrangement in human evolution. Such X-Y homologous regions are relevant to the detection of a role of retroviruses/retroposons in recent evolution in two respects. First, examination of homologous sequences on the X and the Y may reveal new insertions and allow them to be dated. Second, investigation of the relationship between genes within these regions and such elements as may be related to them could uncover a role in gene expression since any recent change will be differentially expressed in the two sexes.

The Xq21.3/Yp11.2 region of homology is of particular interest in relation to human evolution, as it was created by a transposition after the separation of the chimpanzee and human lineages, and 3.5 Mb of the transposed region was subject to a subsequent paracentric inversion (Sargent *et al*., 1996; Mumm *et al*., 1997). We identified two retroposons (HS307 and HS408) within the Xq21.3 region with homology to the SINE-R.C2 element (Kim *et al*., 1999b). It will be important to examine the homologies of these elements within the Yp11.2 region to ascertain whether changes in these or other elements (eg SINE-R.Y-2 or Y-7) might be associated with genes which in turn are relevant to the sex differences that have been described in relation to Homo sapiens specific cognitive functions (Crow *et al*., 1998).

SINE-R11, 14 and 19 have a polypurine tract (PPT) that serves as a primer-binding-site for plus-strand DNA synthesis from retrovirally mediated reverse transcription. This indicates that SINE-R-type elements like human endogenous retroviruses have a potential functional role in the genomic DNA of human chromosomes. According to PCR and phylogenetic analyses, integration of several HERV- K LTR elements have been shown to be human specific (Medstrand and Mager, 1998). We also identified a solitary HERV-K LTR element in a recently proliferated class of LTR elements, this was inserted into Xq26 after the separation of the hominid and great ape lineages (Kim *et al*., submitted). It seems that the LTR elements have been randomly transposed across the chromosome in the course of evolution. Apokov *et al*. (1998) noted that such sequences have the capacity to modify the expression of neighboring genes. These elements may also have contributed to genetic variation associated with various disease. The HERV-K sequences, to which the SINE-R elements are related, have been reported as a candidate autoimmune gene in type I diabetes (Conrad *et al*., 1997). Two HERV-K LTR elements have been detected in the human major histocompatibility complex locus HLA-DQ (Kambhu *et al*., 1990). One LTR element (DQ-LTR3) of the HERV-K family at the HLA-DQB1 locus has been associated with the rheumatoid arthritis (Seidl *et al*., 1999).

ACKNOWLEDGMENTS

We thank Rekha V. Wadekar for her technical assistance. This work was supported by grants from the UK Medical Research Coun-

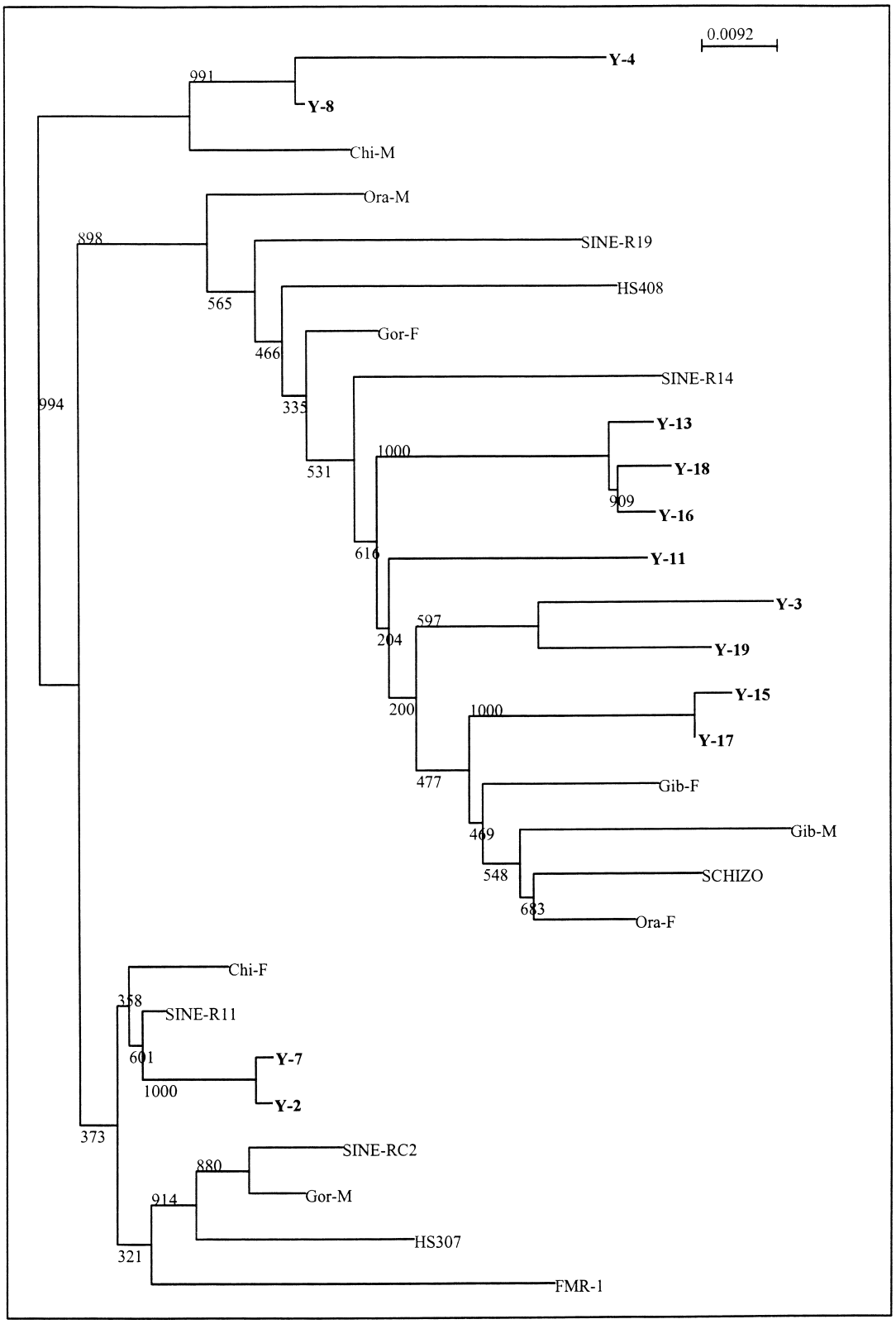

Fig. 2. Phylogenetic tree obtained by neighbor-joining method for the retroposon family. Statistical robustness was tested by bootstrap analysis. The numbers to the left of the nodes of the tree branches represent the number of times out of 1000 trees. The retroposons (Chi-M, Chi-F, Gor-M, Gor-F, Ora-M, Ora-F, Gib-M, Gib-F) from hominoid primates were taken from GenBank database (see Materials and Methods).

cil and the Stanley Foundation.

REFERENCES

Akopov SB, Nikolaev LG, Khil PP, Lebedev YB, Sverdlov ED (1998) Long terminal repeats of human endogenous retrovirus K family (HERV-K) specifically bind host cell nuclear proteins. FEBS Lett 421: 229–233

Altschul SF, Madden TL, Schäffer AA, 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

Baltimore D (1985) Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. Cell 40: 481–482

Carroll MC, Campbell RD, Bentley DR, Porter RR (1984) A molecular map of the human major histocompatibility complex class III region linking complement genes C4, C2 and factor B. Nature 307: 237–241

Conrad B, Weissmahr RN, Boni J, Arcari R, Schupbach J, Mach B (1997) A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. Cell 90: 303–313

Crow TJ, Crow LR, Done DJ, Leask SJ (1998) Relative hand skill predicts academic ability: global deficits at the point of hemispheric indecision. Neuropsychologia 36: 1275–1282

Glaser B, Grutzner F, Taylor F, Schiebel K, Meroni G, Tsiouprа K, Pasantes J, Rietschel W, Toder R, Willman U, Zeitler S, Yen P, Ballabio A, Rappold G, Schempp W (1997) Comparative mapping of Xp22 genes in hominoids evolutionary linear instability of their Y homologues. Chrom Res 5: 167–176

Haltimeier M, Seifarth W, Blusch J, Erfle V, Hehlmann R, Leib-Mosch C (1995) Identification of S71-related human endogenous retroviral sequences with full-length pol genes. Virology 209: 550–560

Jegalian K, Page DC.(1998) A proposed path by which genes common to the X and Y chromosomes evolve to become X inactivated. Nature 394: 776–780

Kambhu S, Falldorf P, Lee JS (1990) Endogenous retroviral long terminal repeats within the HLA-DQ locus. Proc Natl Acad Sci USA 87: 4927–4931

Kazazian HH, Moran JV (1998) The impact of the L1 retrotransposons on the human genome. Nature Genet 19: 19–24

Kelsall DP, Rooke L, Warne D, Bonzyk M, Cullin L, Cox S, West L, Povey S, Spurr NK (1995) Development of a panel of monochromosomal somatic cell hybrids for rapid gene mapping. Ann Hum Genet 59: 233–241

Kim H-S, Chen Y, 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, Takenaka O (1996) Molecular features of the TSPY gene of gibbons and Old World monkeys. Chrom Res 4: 500–506

Kim H-S, Takenaka O, Crow TJ (1999a) Cloning and nucleotide sequence of retroposons specific to hominoid primates derived from an endogenous retrovirus (HERV-K). AIDS Res Hum Retroviruses 15: 595–601

Kim H-S, Wadekar RV, Takenaka O, Winstanley C, Mitsunaga F, Kageyama T, Hyun B-H, Crow TJ (1999b) SINE-R.C2 (a Homo sapiens specific retroposon) is homologous to cDNA from postmortem brain in schizophrenia and to two loci in the Xq21.3/Yp block linked to handedness and psychosis. Am J Med Genet (Neuropsychiat Genet) (in press)

Kobayashi K, Nakahori Y, Miyake M, Matsumura K, Kondo-lida E, Nomura Y, Segawa M, Yoshioka M, Saito K, Osawa M, Hamano K, Sakakihara Y, Nonaka I, Nakagome Y, Kanazawa I, Nakamura Y, Tokunaga K, Toda T (1998) An ancient retrotransposal insertion causes Fukuyama-type congenital muscular dystrophy. Nature 394: 388–392

Laval SH, Dann JC, Butler RJ, Loftus J, Rue J, Leask SJ, Bass N, Comazzi M, Vita A, Nanko S, Shaw S, Peterson P, Shields G, Smith AB, Stewart J, DeLisi LE, Crow TJ (1998) Evidence for linkage to psychosis and cerebral asymmetry (relative hand skill) on the X chromosome. Am J Med Genet (Neuropsychiat Genet) 81: 420–427

Lambson B, Affara NA, Mitchell M, Ferguson-Smith MA (1992) Evolution of DNA sequence homologies between the sex chromosomes in primate species. Genomics 14: 1032–1040

Leib-Mösch C, Haltmeier M, Werner T, Geigl E-M, Brack-Werner R, Francke U, Erfle V, Hehlmann R (1993) Genomic distribution and transcription of solitary HERV-K LTRs. Genomics 18: 261-269

Mager DL, Freeman JD (1995) HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. Virology 213: 395–404

Medstand P, Mager DL (1998) Human-specific integrations of the HERV-K endogenous retrovirus family. J Virol 72: 9782–9787

Mumm S, Molini B, Terrell J, Srivastava A, Schlessinger D (1997) Evolutionary features of the 4 Mb Xq21.3 XY homology region revealed by a map at 6 kb resolution. Genome Res 7: 307–314

Ono M, Yasunaga T, Miyata T, Ushikubo H (1986) Nucleotide sequence of human endogenous retrovirus genome related to the mouse mammary tumor virus genome. J Virol 60: 589–598

Ono M, Kawakami M, Takezawa T (1987) A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res 21: 8725–8737

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425

Sargent CA, Briggs H, Chalmers IJ, Lambson B, Walker E, Affara NA (1996) The sequence organization of Yp/proximal Xq homologous regions of the human sex chromosomes is highly conserved. Genomics 32: 200–209

Schwartz A, Chan DC, Brown LG, Alagappan R, Pettay D, Disteche C, McGillivray B, de la Chapelle A, Page DC (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

Seidl C, Donner H, Petershofen E, Usadel KH, Seifried E, Kaltwasser JP, Badenhoop K (1999) An endogenous retroviral long terminal repeat at the HLA-DQB1 gene locus confers susceptibility to rheumatoid arthritis. Hum Immunol 60: 63–68

Shimamura M, Nikaido M, Ohshima K, Okada N (1998) A SINE acquired a role in signal transduction during evolution. Mol Biol Evol 15: 923–925

Simon M, Haltmeier M, Papakonstantinou G, Werner G, Hehlmann R, Leib-Mosch C (1994) Transcription of HERV-K-related LTRs in human placenta and leukemic cells. Leukemia 8: S12–17

Steinhuber S, Brack M, Hunsmann G, Schwelberger H, Dierich MP, Vogetseder W (1995) Distribution of human endogenous retrovirus HERV-K genomes in humans and different primates. Hum Genet 96: 188–192

Sverdlov ED (1998) Perpetually mobile footprints of ancient infections in human genome. FEBS Lett 428: 1–6

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673–4680

Zhu Z-B, Hsieh S-L, Bentley DR, Campbell RD, Volanakis JE (1992) A variable number of tandem repeats locus within the human complement C2 gene is associated with a retroposon derived from a human endogenous retrovirus. J Exp Med 175: 1783–1787

Zhu Z-B, Jian B, Volanakis JE (1994) Ancestry of SINE-R.C2 a human-specific retroposon. Hum Genet 93: 545–551

(Received December 4, 1998 / Accepted June 10, 1999)