Complete Nucleotide Sequences of ALV-Related Endogenous Retroviruses Available from the Draft Chicken Genome Sequence

(chicken / endogenous retroviruses / ALV genus / expression)

L. BORISENKO, A. V. RYNDITCH

Department of Molecular Oncogenetics, Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences, Kiev, Ukraine

Abstract. Complete nucleotide sequences of chicken endogenous retroviruses belonging to E33/E51 and EAV-0 groups have been analysed on the basis of the recently available draft genome sequence of red jungle fowl (*Gallus gallus*), the progenitor of domestic chicken (*G.g. domesticus*). It was shown that all these proviruses have deletions in the SU-coding domain of the *env* gene, involved in receptor recognition, whereas *gag* and *pol* genes appear to be intact. Phylogenetic analysis demonstrated that E33/E51 and EAV-0 groups are related to the ALV genus. An analysis of expression using chicken EST databases showed that these proviruses are transcriptionally active.

Endogenous proviruses are copies of exogenous retroviruses integrated into the host germ line cells. They are inherited vertically as Mendelian genes and are known to be widespread within the genomes of all vertebrates (Herniou et al., 1998). Endogenous retroviruses of the domestic chicken remain described incompletely, in spite of intensive studies.

All known chicken endogenous retroviruses can be subdivided into three families: (1) the ev loci (belonging to the ALV genus), (2) the EAV family with members E51, E33, EAV-HP, EAV-0 and ART-CH (genus unknown) and (3) the human endogenous retrovirus type I (HERV-I)-related retroviruses (belonging to the MLV genus) (see Borisenko, 2003, for review). The ev loci are specific for the domestic chicken and its wild relative, the red jungle fowl, while the EAV family is present in all Gallus species (Boyce-Jacino et al., 1992). In contrast, HERV-I-related proviruses have been found in the genomes of all vertebrates (Martin et al., 1997). Such a distribution suggests that the EAV family is more ancient than the ev loci. The lack of complete genomic sequence for EAV proviruses (E51, E33, EAV-0) has made it difficult to determine the origin of this heterogeneous group. The recent sequencing of the red jungle fowl genome has provided an opportunity to clarify this question. Here we present a description of the genomic structure of EAV proviruses, their phylogenetic analysis and their patterns of expression.

Identification of proviruses within the draft chicken genome sequence was made using the BLAT program (http://genome.ucsc.edu) and GenBank sequences (M31063, X59844 – for EAV-0; M95189 – for E51; M95190 – for E33; L25262 – for ART-CH) as queries. Multiple sequence alignments were carried out using Clustal W (Thompson et al., 1994) and open reading frames (ORF) were detected using ORFfinder (http://www.ncbi.nlm.nih.gov).

E33 and E51 (Boyce-Jacino et al., 1992) searches identified sequences with an unusual degree of homology. Because they are more closely related to each other than to other groups of chicken endogenous retroviruses, we united them in a separate group, E33/E51, to distinguish them from the other EAV elements – EAV-HP and EAV-0.

To estimate the degree of within-group divergence, several E33/E51 proviruses with full-length long terminal repeats (LTRs) were used to calculate similarity. Table 1 shows that E33/E51 sequences have a similarity range of 75–99%. This value is high compared to that of other chicken endogenous retrovirus groups (Table 2). For example, EAV-0 group members range only from 93% to 100% in similarity, between proviruses localized on different chromosomes (data not shown).

All E33/E51 proviruses contain multiple deletions in the SU-coding domain of their *env* genes. Interestingly, proviruses localized on different chromosomes have similar breakpoints, suggesting they may be offspring of a provirus progenitor. Figure 1 shows the locations of gaps in the E33/E51 SU-coding domain in comparison to RSV SU.

No intact EAV-0 proviruses were detected in the chicken genome sequence. The deletion in the EAV-0 *env* gene spans almost the entire SU-coding domain as previously reported (Boyce-Jacino et al., 1992).

Because the SU-coding domain is involved in receptor recognition and is required for virus entry, E33/E51 and EAV-0 are unable to produce infectious virions.

Received May 17, 2004.

Corresponding author: Leonid Borisenko, Department of Molecular Oncogenetics, Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences, Kiev, Ukraine. Tel.: 038-044-2663498; fax: 038-044-2663498; e-mail: rynditch@imbg.org.ua.

Abbreviations: ALV – avian leukosis virus, ART-CH – avian retrotransposon from the chicken genome, EAV – endogenous avian retrovirus family, HERV-I – human endogenous retrovirus type I, MLV – murine leukaemia virus.

Table 1. LTR sequence similarity among E33/E51 proviruses from the draft chicken genome sequence

LTRs[1]	% similarity to[2]:						
	E33	E51	seq1	seq2	seq3	seq4	seq5
E51	78.5						
seq1	94.4	77.9					
seq2	85.0	76.5	84.9				
seq3	78.7	99.2	78.2	76.7			
seq4	77.7	89.4	75.9	74.2	89.6		
seq5	77.9	89.1	76.9	74.4	89.4	92.2	
seq6[3]	79.0	93.2	78.3	77.5	93.5	91.0	91.5

[1]GenBank accession numbers of sequences used in the alignment: E-33 - M95190, E51 - M95189. Positions of LTRs within the draft chicken genome sequence: seq1 - 150476453-150476824 on chromosome 1("-" strand), seq2 - 22045601-22045982 on chromosome 6(-), seq3 - 971197-971557 on chromosome Z(-), seq4 - 115042-115408 on chromosome 28(-), seq5 - 55552889-55553256 on chromosome 2(-), seq6 - 69175051-69175418 on chromosome 1(+).
[2]sequences were aligned using program CLUSTAL W (Thompson et al., 1994) and percent similarities as [1-*P* distances]100 were calculated.
[3]sequence between LTRs demonstrates homology to ART-CH.

Table 2. DNA sequence similarity among chicken endogenous retroviruses

Retrovirus[1]	% similarity to[2]:		
	ev-1	EAV-HP	E33/E51
EAV-HP	41.9/50.4/59.7/52.0		
E33/E51	47.0/50.6/60.2/50.1	45.3/54.9/64.2/57.0	
EAV-0	53.8/53.5/62.7/NI	51.9/49.8/62.5/NI	50.7/52.0/63.2/NI

[1]GenBank accession numbers of sequences used in the alignment: ev-1 - AY013303; EAV-HP - AJ292967. Positions of proviruses within the draft chicken genome sequence: E33/E51 - 115042-122630 on chromosome 28 ("-" strand), EAV-0 - 6373152-6379336 on chromosome 2(-).
[2]calculation of percent similarity - as in Table 1. First number - LTR similarity, second number - *gag* gene similarity, third number - *pol* gene similarity, fourth number - *env* gene similarity. NI - not identified since large deletions made the comparison invalid.

The E33/E51 and EAV-0 proviruses share structural features with other chicken endogenous retroviruses and with RSV (Table 3). In addition, all have similar primer binding sites for tRNA-Trp and polypurine tracts agggaggggga.

ART-CH is another retroelement related to the EAV family. The U3 region of ART-CH LTR is more than 92% identical to that of E51 (Nikiforov and Gudkov, 1994). However, ART-CH with full-length LTRs similarity to the E33/E51 group were also found in the chicken genome sequence (see for example seq6 in Table 1). This finding confirms assumptions about LTR domain shuffling resulting from recombination between different chicken retroviruses (Boyce-Jacino et al., 1992) and suggests that ART-CH is a chimeric product of such recombination (Nikiforov and Gudkov, 1994).

To elucidate the phylogenetic relationships of the E33/E51 and EAV-0 groups, a phylogenetic tree was constructed using PHYLIP 3.5 (Felsenstein, 1989). This was based on the *pol* gene because it is the most conserved among the retroid elements for which *pol*-based trees have been constructed before (Xiong and Eickbush, 1990). N-terminal amino acid residues (from domain 1 to 5 as indicated by Xiong and Eickbush, 1990) from 17 retroelements were used in the analysis: 1) chicken endogenous retroviruses: ev-1, EAV-HP, E33/E51, EAV-0, 2) retroviruses, mostly exogenous, representing all seven retroviral genera: ALV genus (Rous sarcoma virus, RSV), MLV-related retroviruses

This is likely the result of a long history of retrovirus-host co-evolution providing a selection for non-pathogenic proviruses. Loss of part of the *env* gene is a common phenomena among ancient endogenous retroelments (Coffin et al., 1997).

The *gag* and *pol* genes of E33/E51 and EAV-0 are generally intact, although some are interrupted by in-frame stop-codons. The non-deleted *gag-pol* regions probably encode an active reverse transcriptase, as has been demonstrated for EAV-0 (Weissnahr et al., 1997).

Table 3. Comparison of chicken endogenous retroviruses and RSV

Retro-virus[1]	LTR size (bp)				5' non-translated region size (bp)	*gag* size (bp)	*pol* size (bp)	*env* size (bp)	3' non-translated region size (bp)	Retro-virus size (bp)
	total	U3	R	U5						
RSV	335	234	21	80	270	2105	2709	1786	222	8046[2]
ev-1	274	174	21	79	280	2105	2391	1833	165	7525
E33/E51	362	242	22	98	315	2260	2700	1385	181	7588
EAV-0	240	141	21	78	241	2240	2692	580	82	6183
EAV-HP	314	175	17	122	151	2015	2580	1750	180	7120

[1]GenBank accession numbers of sequences used in the alignment: RSV - AF033808, others - as in Table 2.
[2]size without 1580 bp *src* gene.

```
RSV      GTTCACTTACTCGAGCAGCCAGGGAACCTTTGGATTACATGGGCCAACCGTACAGGCCAA  5308
E33/E51  ..A...C.TG.AC.A......CAC..TG.....G.C..G.....G..T.TC.....G.GG

RSV      ACGGATTTCTGCCTCTCTACACAGTCAGCCACCTCCCCTTTTCAAACATGTTTGATAGGT  5368
E33/E51  ..A..C.....TT.GGGCCTT.C...C..T.....T.....C.GT..C.....A......

RSV      ATCCCGTCTCCTATTTCCGAAGGTGATTTTAAGGGATATGTTTCTGATACAAATTGCTCC  5428
E33/E51  T.G..AAA.TA.-CAATT.....A.--.........C..AC----...-...C..CAGTG

RSV      ACTGTGGGAACTGACCGGTTAGTCTTG-TCAGCCAGCATTACCGGCGGCCCTGACAACAG  5487
E33/E51  TG...T....TGA.A.A.ACGC.GC.AC...AA.G..G.----.T.T.ATT...TC.TTA

RSV      CACCACCCTCACTTATCGAAAGGTTTCATGCCTGCTGTTAAAGCTGAACGTCTCCATGTG  5547
E33/E51  A....TA.C.T.C.C.G.G.C-------------.CCCA.G.AT..G.TA.T.TAGG..C

RSV      GGATGAGCCACCTGAACTGCAGCTGCTAGGTTCCCAGTCTCTCCCTAACGTTACTAACAT  5607
E33/E51  TC.--.ATG.T.A.G.A..G.A.AA.AC.--.A.GT..G..A..T.TGGT.C.A.G-TGC

RSV      TACTCAGGTCTCTGGCGTGGCCGGGGGATGTGTATATTTCGCCCCAAGGGCCACTGGCCT  5667
E33/E51  ..TA...AGAA.GAT...A.T-A.A.TC...CACA....T.ATGGG.ATTTTGA...--G

RSV      GTTTTTAGGTTGGTCTAAACAAGGTCTCTCGCGGTTCCTCCTCCGTCACCCCTTTACCTC  5727
E33/E51  AC.GG.G..G...---.GG..GAA.TG.GT.A---C.T.ATAGT.AG.TGGGG.A.TAAT

RSV      CACCTCTAACTCCACGGAACCGTTCACGGTGGTGACAGCGGATAGACACAATCTTTTTAT  5787
E33/E51  G...CTCG----T.TAA....T.-----A..C.A..T.ATC..G...G----------G.

RSV      GGGGAGT—GAGTACTGTGGTGCATATGGCTACAGATTTTGGGAAATATATAACTGCTCA-  5845
E33/E51  ..T....CC..TG.ACA...A.AG.T.TT.G.T..C.GG..C.T.T.G.GGT----T...

RSV      CAGACTAGGAATACTTACCGCTGTGGAGACGTGGGAGGTACTGGCCTCCCTGAAACCTGG  5905
E33/E51  ..A.GA.C...-...CGTTAT.A.AA..------.G.A.T..------T....---T...

RSV      TGCAGAGGAAAAGGAGGTATATGGGTTAATCAATCAAAGGAAATTAATGAGACAGAGCCG  5965
E33/E51  ..T...TC.........G.A....-.C.GAGGGA..C..A..-----..G....ATATGC

RSV      TTCAGTTTTACTGCGAACTGTACTGGCAGTAATTTGGGTAATGTCAGCGGATGTTGCGGA  6025
E33/E51  .CTG.G.GC.G..GT..T..C..-A....--.A.G.AAC...-.AT..AT...-------

RSV      GAACCAATCACGATTCTCCCACTAGGGGCATGGATCGACAGTAC-GCAAGGTAGTTTCAC  6084
E33/E51  -----------.G...A..TTTAG.A..A.G.T....GAG...TT.TGGAA..A.AAG..

RSV      TAAACCAAAAGCGCTACCACCCGCAATTTTCCTCATTTGTGGGGATCGCGCATGGCAAGG  6144
E33/E51  .G---.T.....A..C..C..A.G...C...T.G........CA..A.G..........

RSV      AATTCCCAGTCGTCCGGTAGGGGGCCCCTGCTATTTAGGCAAGCTTACCATGTTAGCACC  6204
E33/E51  TG.C..AGC.AA.T.TC.G..A..T..G..T..C.....G...T.A..G.....T..T..

RSV      CAACCATACAGATATTCTCAAAATACTTGCTAATTCGT-CGCGGACAGGTATAAGACGTA  6263
E33/E51  A..T..C....GATGG..T..T...TC.CGC.G..TACATC.TCG....CGCC-.T..C.

RSV      AACGA  6268
E33/E51  GT...
```

Fig. 1. Nucleotide sequence comparison of RSV and E33/E51 SU-coding domains. Sequences were aligned using program CLUSTAL W (Thompson et al., 1994); positions of nucleotides correspond to GenBank RSV sequence AF033808; localization of E33/E51 within the draft chicken genome sequence: 55545429-55553256 on chromosome 2 ("-" strand); dot indicates nucleotide identity with RSV; dash indicates gap introduced during sequence alignment.

```
CERV-I          LKDGLLE--PCMSPFNTPILPVRKP------DGSYRLVQDLRKINEIVQKRHPAVPNPYT
HERV-I          INDGLLE--PCMSPYNTPILPVKKS------DGSYRLVKDLRAINQTVQTTNPVVPNPYT
MuLV            LDQGILV--PCQSPWNTPLLPVKKPG-----TNDYRPVQDLREVNKRVEDIHPTVPNPYN
HSRV            LKQGVLT--PQNSTMNTPVYPVPKP------DGRWRMVLDYREVNKTIPLTAAQNQHSAG
HTLV-1          LEAGHIE--PYTGPGNNPVFPVKKA------NGTWRFIHDLRATNSLTIDLSSSSPGPP-
BLV             LEAGYIS--PWDGPGNNPVFPVRKP------NGAWRFVHDLRATNALTKPIPALSPGPP-
E33/E51         FQLGHIE--PSLSQWNTPIFVIQKR------SGTFRLLHDLRAVNAQLVPFGAVQQGGP-
EAV-0           LEAGHIE--PSLSRWNTPIFVIRKP------SGSFRLLHDLRAVNAQLVQFGPVQQGGP-
RSV             LQLGHIE--PSLSCWNTPVFVIRKA------SGSYRLLHDLRAVNAKLVPFGAVQQGAP-
ev-1            LQLGHIE--PSLSCWNTPVFVIRKA------SGSYRLLHDLRAVNAKLVPFGAVQQGAP-
EAV-HP          LRLGHIE--PSLSRWNTPVFVIQKK------SGAFRLLHDLRAVNSQLIPFGVVQQGAP-
MPMV            LEAGHIT--ESSSPWNTPIFVIKKK------SGKWRLLQDLRAVNATMVLMGALQPGLP-
SRV-1           LEAGHIT--ESNSPWNTPIFVIKKK------SGKWRLLQDLRAVNATMVLMGALQPGLP-
MMTV            LQLGHLE--ESNSPWNTPVFVIKKK------SGKWRLLQDLRAVNATMHDMGALQPGLP-
HIV-1           --EGKISKIGPENPYNTPVFAIKKKD-----STKWRKLVDFRELNKRTQDFWEVQLGIP-
FIV             --EGKVKRADSNNPWNTPVFAIKKK------SGKWRMLIDFRELNKLTEKGAEVQLGLP-
gypsy           LKDGIIR--PSRSPYNSPTWVVDKKGTDAFGNPNKRLVIDFRKLNEKTIPDRYPMPSIPM

CERV-I          LMSKIPNENRW-FSVIDLKDAFWSIPLDHESRDIFAFEWEDPE-SGRKQQYRWTVLPQGF
HERV-I          ILSKIPYNHQW-FTVIDLKDAFWACPLAEESRDTFAFEWEDPQ-LGXKQWYQWTVLPQGF
MuLV            LLSGLPPSHQW-YTVLDLKDAFFCLRLHPTSQPLFAFEWRDPE-MGISGQLTWTRLPQGF
HSRV            ILATIVR-QKY-KTTLDLANGFWAHPITPESYWLTAFTWQG-------KQYCWTRLPQGF
HTLV-1          DLSSLPTTLAH-LQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGF
BLV             DLTAIPTHPPH-IICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGF
E33/E51         ILSAIPKEWP--LVVVDLKDCFFSIPLTEEDREAFAFTVPTLNNLGPTERFQWRVLLQGM
EAV-0           SLAAVPRGWP--LVVIDLKDCFFSIPLAEQDREAFAFTVPVRNNQGPAQRFQWKVLPQGM
RSV             VLSALPRGWP--LMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGM
ev-1            VLSALPRGWP--LMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGM
EAV-HP          VLSAVPEEWE--VTAIDLKDCFFSIPLAEQDREAFAFTVPVSNNQRPTQRYQWRVLPQGM
MPMV            SPVAIPQGYL--KIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRPQWKVLPQGM
SRV-1           SPVAIPQGYL--KIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQRM
MMTV            SPVAVPKGWE--IIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGM
HIV-1           HPAGLKKKKS--VTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGW
FIV             HPAGLQIKKQ--VTVLDIGDAYFTIPLDPDYAPYTAFTLPRKNNAGPGRRFVWCSLPQGW
gypsy           ILANLGKAKF--FTTLDLKSGYHQIYLAEHDREKTSFSVNG-------GKYEFCRLPFGL

CERV-I          TESPNLF------------------YFLQYVDDL
HERV-I          MDSPNLFGQILEQVLDKVSVPK-QLCLLQYVDD-
MuLV            KNSPTLFDEALHRDLADFRIQHPDLILLQYVDDL
HSRV            LNSPALFTADVVDLLKEIP------NVQVYVDDI
HTLV-1          KNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDI
BLV             INSPALFERALQEPLRQVSAAFSQSLLVSYMDDI
E33/E51         ACSPTICQLVVGRVLEPIRRDFPRYILVHYMDDL
EAV-0           ACSPTICQLVVNTIIAPVRRDMPDCQIVHYMDDL
RSV             TCSPTICQLVVGQVLEPLRLKHPSLCMLHYMDDL
ev-1            TCSPTICQLVVGQVLEPLRLKHPSLRMLHYMDDL
EAV-HP          ACSPTICQMVVGKILGPLHHTSEASEILHYMDDL
MPMV            ANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDI
SRV-1           ANSPTLCQKYVATAIHKVRHAWKQMYIIHYMDDI
MMTV            KNSPTLCQKFVDKAILTVRDKYQDSYIVHYMDDI
HIV-1           KGSPAIFQSSMTKILEPFRKQNPDIVIYQYMDDL
FIV             ILSPLIYQSTLDNIIQPFIRQNPQLDIYQYMDDI
gypsy           RNASSIFQRALDDVLREQIG----KICYVYVDDV
```

Fig. 2. Partial RT protein sequence alignment used for construction of a phylogenetic tree. Data on exogenous retrovirus sequences and *gypsy* are from Xiong and Eickbush (1990); gaps are indicated by dashed lines; unknown amino acids appeared as a result of unknown nucleotide insertion marked as „X“.

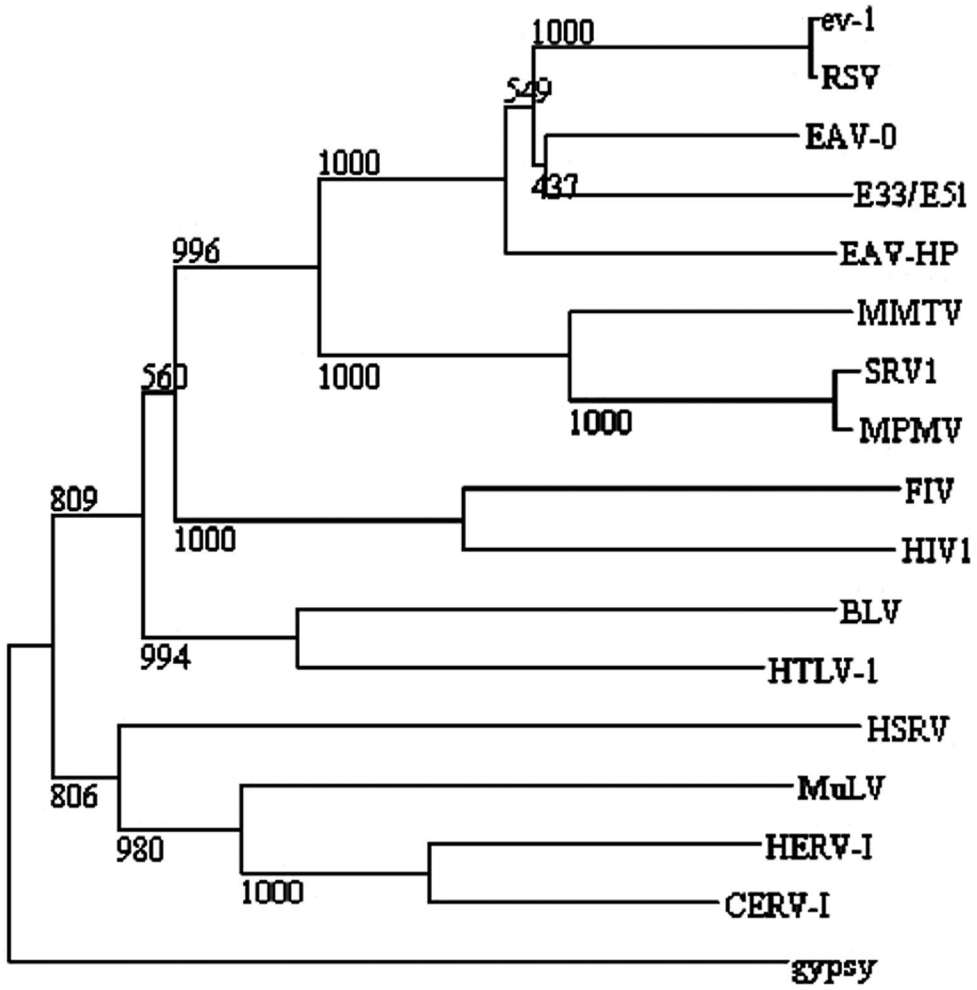

Fig. 3. Phylogenetic neighbour-joining tree of retroviral RT amino acid sequences. The tree was rooted on *gypsy* LTR retrotransposon sequences; numbers at nodes indicate the percent recovery of these nodes per 1000 bootstrap replicates.

(murine leukaemia virus, MLV; human endogenous retroviruses type I, HERV-I; chicken HERV-I-related retrovirus, CERV-I), type D viruses (Mason-Pfizer monkey virus, MPMV; simian AIDS virus type 1, SRV-1), mammalian type B viruses (mouse mammary tumour virus, MMTV), lentiviruses (human immunodeficiency virus 1, HIV-1; feline immunodeficiency virus, FIV), spumaviruses (human spumaretrovirus, HSRV) and HTLV-related viruses (human T-cell leukaemia virus type 1, HTLV-1; bovine leukaemia virus, BLV) and 3) drosophila gypsy LTR-retrotransposon (Fig. 2).

Phylogenetic analysis (Fig. 3) indicates that E33/E51, EAV-0 and EAV-HP do not cluster in one group, although EAV-0 and E33/E51 are more closely related to each other. All these proviruses are more distantly related to RSV than to ev loci. Therefore, E33/E51, EAV-0, EAV-HP and ev loci belong to the ALV genus and the only chicken endogenous retrovirus that is not a member of this genus is CERV-I, which is related to HERV-I and MLV. The description of CERV-I has been presented elsewhere (Borisenko and Rynditch, 2003).

It is likely that the EAV family, consisting of subfamilies EAV-HP, E33/E51 and EAV-0 (Sacco et al., 2000), is artificial, with different, but related, retroviruses belonging to the ALV genus. Regions of homology shared by these elements may be due to recombination events. We now conclude that all chicken endogenous retroviruses infect the host genome separately and probably at different times. Experimental data confirm this point of view that the E33/E51 group is older than EAV-0 (Boyce-Jacino et al., 1992).

We investigated the expression of E33/E51 and EAV-0 proviruses using three chicken EST databases: (1) a bursal library (Abdrakhamanov et al., 2000; Buerstedde et al., 2002; http://swallow.gsf.de/dt40Est.html), (2) the BBSRC libraries (Boardman et al., 2002; http://www.chick.umist.ac.uk) and (3) Delaware Biotechnology Institute database (Tirunagaru et al., 2000; http://www.chickest.udel.edu). All EAV proviruses are transcriptionally active both in adult and embryonic tissues. We also detected expression in chicken embryo fibroblasts using RT-PCR (data not shown).

Lastly, as previously suggested for EAV-HP (Sacco et al 2000), there are tissue specificities in expression: the majority of EAV-HP cDNA clones derived from adult pancreas and no clones were present in the liver. In contrast, the opposite situation has been observed for ART-CH and EAV-0.

The number of gene-specific transcripts also differs among proviruses. For example, the majority of *pol*-transcripts derive from EAV-0, which, along with structurally complete ev loci, may be the main provider of reverse transcriptase in the chicken genome.

The sequencing of the chicken genome has made possible the determination of the complete genomic structures of endogenous retroviruses. Further directions will include analysis of the patterns of proviral localization.

References

Abdrakhmanov, I., Lodygin, D., Geroth, P., Arakawa, H., Law, A., Plachy, J., Korn, B., Buerstedde, J.-M. (2000) Large database of chicken bursal ESTs as a resource for the analysis of vertebrate gene function. *Genome Res*. **10,** 2062-2069.

Boardman, P. E., Sanz-Ezquerro, J., Overton, I. M., Burt, D. W., Bosch, E., Fong, W. T., Tickle, C., Brown, W. R., Wilson, S. A., Hubbard, S. J. (2002) A comprehensive collection of chicken cDNAs. *Curr. Biol*. **22,** 1965-1969.

Borisenko, L. (2003) Avian endogenous retroviruses. *Folia Biol. (Praha)* **49,** 177-182.

Borisenko, L., Rynditch, A. V. (2003) Novel endogenous retrovirus from the chicken genome. *Biopolimers and Cell* **19,** 270-273. (in Russian)

Boyce-Jacino, M., O'Donoghue, K., Faras, A. (1992) Multiple complex families of endogenous retroviruses are highly conserved in the genus Gallus. *J. Virol.* **66,** 4919-4929.

Buerstedde, J.-M., Arakawa, H., Watahiki, A., Carninci, P. P., Yoshihide Hayashizaki, Y., Korn, B., Plachy, J. (2002) The DT40 web site: sampling and connecting the genes of a B cell line. *Nucleic Acids Res.* **30,** 230-231.

Coffin, J., Hughes, S., Varmus, H. (1997) *Retroviruses.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Felsenstein, J. (1989) PHYLIP – Phylogeny Inference Package (version 3.2). *Cladistics* **5,** 164-166.

Herniou, E., Martin, J., Miller, K., Cook, J., Wilkinson, M., Tristem, M. (1998) Retroviral diversity and distribution in vertebrates. *J. Virol*. **72,** 5955-5966.

Martin, J., Herniou, E., Cook, J., O'Neill, R., Tristem, M. (1997) Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates. *J. Virol.* **71,** 437-443.

Nikiforov, M., Gudkov, A. (1994) ART-CH: a VL30 in chickens*? J. Virol.* **68,** 846-853.

Sacco, M., Flannery, D., Howes, K., Venugopal, K. (2000) Avian endogenous retrovirus EAV-HP shares regions of identity with avian leukosis virus subgroup J and the avian retrotransposon ART-CH. *J. Virol.* **74,** 1296-1306.

Tirunagaru, V. G., Sofer, L., Cui, J., Burnside, J. (2000) An expressed sequence tag database of T-cell-enriched activated chicken splenocytes: sequence analysis of 5251 clones. *Genomics* **66,** 144-151.

Thompson, J. D., Higgins, D. G., Gibson, T. J. (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.

Weissmahr, R., Schupbach, J., Boni, J. (1997) Reverse transcriptase activity in chicken embryo fibroblast culture supernatants is associated with particles containing endogenous avian retrovirus EAV-0 RNA. *J. Virol*. **71,** 3005-3012.

Xiong, Y., Eickbuch, T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9,** 3353-3362.