JOURNAL OF VIROLOGY, Aug. 1992, p. 4919–4929
0022-538X/92/084919-11$02.00/0

Vol. 66, No. 8

Multiple Complex Families of Endogenous Retroviruses Are Highly Conserved in the Genus *Gallus*

M. T. BOYCE-JACINO,[1]* K. O'DONOGHUE,[1] AND A. J. FARAS[1,2]

Institute of Human Genetics[1] and Department of Microbiology,[2] University of Minnesota, Minneapolis, Minnesota 55455

Received 19 December 1991/Accepted 13 May 1992

We have analyzed the genome of the domestic chicken for the presence of genetic sequences related to the envelope protein-encoding genes of avian sarcoma/leukosis retroviruses to determine the organization, structure, potential functionality, and distribution of such sequences. We have previously identified in the genus *Gallus* an extensive group of endogenous avian retroviruses termed EAV-0. Southern blot and sequence analysis presented here of EAV-0 elements revealed that the majority of the EAV-0 elements in the domestic chicken genome have large deletions in their *env* genes. Screening of a line 0 chicken genomic DNA library for potential full-length *env* gene-containing endogenous elements yielded three provirus clones of a previously unrecognized group of endogenous retroviruses. These three clones, E13, E33, and E51, are more closely related to each other (80% or more sequence identity) than to other avian retroviruses (70% or less sequence identity). The E13 element has a large deletion in *env*, but the E51 element has full-length and highly divergent SU- and TM-coding domains. Complete sequence analysis of the E51 *env* gene region revealed a defective SU-coding domain and an intact TM-coding domain. Sequence analysis of the E51, E33, and E13 3′ termini revealed highly distinctive long terminal repeats of approximately 360 bp which appear to be the products, in part, of long terminal repeat domain shuffling. Hybridization analysis with E51 and E33 *env* gene probes indicated that they are members of an extensive group of elements present in all *Gallus* species, and at least one element, E51, could be shown by polymerase chain reaction amplification and direct sequencing to have integrated prior to *Gallus* speciation.

We have described previously the isolation and characterization of a group of endogenous proviruslike elements which are present at approximately 50 copies per haploid genome in the avian genus *Gallus* (10, 11). This group of proviruses, termed EAV (endogenous avian retrovirus)-0, is highly conserved in the genus *Gallus* and related to endogenous elements in other avian genera (26). Hybridization and sequence analysis showed that EAV-0 elements are distinct from but distantly related to the RAV (Rous-associated virus)-0 family of avian endogenous retroviruses. The EAV-0 elements are more broadly distributed in *Gallus* species than are RAV-0 elements, and at least some integration events predate the divergence of *Gallus gallus* from *Gallus sonerati* (26), indicating that EAV-0 elements are more evolutionarily ancient than RAV-0 elements. The EAV-0 elements are also distinguished by unusually short but functional long terminal repeats (LTRs) (243 bp), also consistent with their being distinct from the avian sarcoma/leukosis retrovirus (ASLV) group (2). A genomic DNA library of the line 0 domestic chicken (*Gallus gallus domesticus*) was screened to determine whether any of the EAV-0 elements or other related but uncharacterized proviral elements in the avian genome contained full-length and potentially functional *env* genes. The *env* genes can be sources of genetic diversity through recombination with exogenous viruses, and their protein products can act as immunologically functional cell surface proteins.

We have shown previously that EAV-0 elements are transcriptionally active during embryogenesis of the domesticated chicken (*G. gallus domesticus*) line 0 (2). Unexpectedly, the largest EAV-0 proviral transcript detected, which hybridized to *gag*, *pol*, *env*, and LTR region probes, was only 5.3 kb, approximately 1.2 kb shorter than a predicted full-length genomic RNA (having intact *gag*, *pol*, and *env* genes) on the basis of collinearity with RAV-0. Here we present an analysis of the deletion breakpoints in several independently isolated EAV-0 clones, which shows that each of the EAV-0 proviruses has a unique deletion in *env*. The uniqueness of each deletion indicates that the EAV-0 proviruses are the products of multiple independent integration events. To determine whether any intact *env* gene-related sequences were present in the chicken genome, we screened all clones from a line 0 genomic DNA library which hybridized to Rous sarcoma virus (RSV) and EAV-0 *env* probes under low-stringency conditions. This screening yielded three clones (E13, E33, and E51) with potentially intact *env* genes. Sequence analysis of these elements indicated that E13 has a large deletion in *env* analogous to that seen in EAV-0 elements. Sequencing of the entire E51 *env* gene indicates that this provirus has an apparently full-length *env* gene with multiple frame shifts in the SU-coding domain. The LTRs of the three clones are long (360 bp for E51) and distinct from any previously characterized retroviral LTRs. The R region of the E51 and E33 LTRs, however, is nearly identical to that of EAV-0 LTRs. Hybridization analysis with *env* region probes from E51 and E33 shows that these elements are members of a highly conserved group present in all *Gallus* species.

The identification of a group of proviruses distinct even from the EAV-0 elements indicates that there is a much greater complexity to avian endogenous retroviruses than previously assumed and that the avian system is analogous in complexity to that of the mouse. Mouse endogenous retroviruses have been well characterized and can be broadly grouped as either closely related to exogenous

* Corresponding author.

viruses (endogenous murine leukemia virus [MLV] and mouse mammary tumor virus [MMTV]) or retroviruslike but unrelated to known exogenous viruses (murine retrovirus-related sequences, intracisternal A-type particles [IAP], and VL30s). The MLV-related endogenous retroviruses of mice are distributed in a fashion analogous to that of the avian RAV-0 endogenous elements; most mouse strains and species have one or none of the different ecotropic MLV loci known (17). The xenotropic MLV endogenous proviruses occur with greater frequency in mouse strains and species (15), but several feral mouse species have been shown by hybridization to lack xenotropic virus-related sequences (20). The polytropic MCF proviruses, which are distinct from xenotropic viruses in their *env* genes (22), are present in many mouse strains but are also absent in some wild mice (21). The MMTV proviruses are also present at a few or no copies in the genomes of various mouse strains (7) and absent in most feral mice species (3, 4). The family of elements in mouse most analogous in distribution to EAV-0 is the murine retrovirus-related sequence family, which is present at 50 to 100 complete copies, and 500 to 1,000 solo LTRs per haploid genome in all mouse strains tested (28). The IAP elements and VL30s elements are two families of much more ancient, and defective, retroviruslike elements in mice; both IAPs and VL30s are present in all species of mice analyzed and in some other, closely related mammals such rats, suggesting that they entered the mouse germ line before rodent species divergence (16, 23, 24).

Analysis of the E51 *env* and LTR sequences presented here indicates that these retroviruses share a common lineage with EAV-0 and ASLVs but are distinct. The E51 and EAV-0 elements are unlike RAV-0 elements or the endogenous MLV and MMTV elements in that they have no known closely related exogenous counterpart. The E51 and EAV-0 elements, however, are structurally more like endogenous retroviruses than IAP or VL30s elements, suggesting that they may represent intermediates in the evolution of extant functional avian retroviruses. We also report the results of polymerase chain reaction (PCR) amplification and direct sequence analysis that indicate that the E51 element integrated prior to divergence of the species *G. varius* and *G. gallus*.

MATERIALS AND METHODS

EAV provirus subclones and RSV clones. The EAV-0 clone used in sequence analysis and as a probe is derived from the λ47.1 chicken genomic DNA library clone 47.1.2 previously described (2). A 1.2-kb deletion subclone of 47.1.2 used for the Southern blot analysis contains the 3′ half of the IN-coding domain and all of the *env* gene sequences (deleted in the SU-coding domain). The 1.2-kb EAV-0 *env* probe clone also lacks LTR and cellular flanking sequences. Because of the conservation of the 5′ *Eco*RI site of this clone, when it is hybridized to *Eco*RI-digested genomic DNA, it detects *env*-cellular junction fragments.

The lambda genomic DNA library clones were isolated by low-stringency plaque lift hybridization analysis as previously described (2). The E51 subclone used for hybridization and sequence analysis was derived from L-charon 30 genomic chicken DNA library clone E51. A 6-kb *pol-env*-LTR piece was subcloned into pGEM, and a 1.4-kb *Bam*HI-*Pst*I *pol* region fragment was used as a probe. Two overlapping, *env*-hybridizing fragments from lambda clone E13, a 1.5-kb *Hin*dIII-*Hin*dIII fragment and a 2.5-kb *Eco*RI-*Eco*RI fragment, were subcloned for sequence analysis and use as probes. A 3.5-kb *Bam*HI-*Hin*dIII *env*-hybridizing fragment of lambda clone E33 was also subcloned for sequence analysis. A 1.2-kb region of E33 which amplified in PCR with the EAV-0 *env*-derived oligomers was also cloned into pGEM.

Plasmid pRSV-*env* was generated by subcloning a 1.7-kb *Xho*I-to-*Xho*I *env* fragment of pATV8 (RSV) Prague C (PRC) bases 5258 to 6983 (18) into pUC17. This clone was used as the complete RSV *env* probe. A 900-bp *Eco*RI-*Xho*I subfragment of the pRSV-*env*, comprising most of the gp85-coding domain (RSV PrC bases 5258 to 6144), was subcloned as pRSV-*env*5′ and was used as the RSV 5′ *env* probe.

Avian genomic DNAs. Blood samples from *G. gallus domesticus* line 11 were obtained from Spafas, Inc. Blood samples from line 0 and *G. sonerati* (green jungle fowl) were generously provided by Lyman Crittenden, U.S. Department of Agriculture Regional Poultry Research Center, East Lansing, Mich. Blood samples from *G. varius* (grey jungle fowl) were generously provided by the San Antonio Zoo and the San Diego Zoo. DNA was prepared from whole blood and digested with restriction enzymes for Southern blot analysis as previously described (26).

DNAs from the other avian species were isolated from embryonated eggs as previously described (10). The eggs were obtained from SPAFAS (Peking duck) or from regional breeders identified by the U.S. Department of Agriculture (Japanese quail, bobwhite quail, peafowl, ringneck pheasant, and red jungle fowl). Red jungle fowl (*G. gallus*) is the feral progenitor of domestic chickens (*G. gallus domesticus*).

Sequence analysis. For sequencing the 3′ end of the EAV-0 provirus, the 3.2-kb *Eco*RI-to-*Eco*RI fragment from L47.1-ev$^-$-2 was cloned into M13mp9. A series of overlapping deletion clones was generated by using the deletion cloning method of Dale et al. (9). DNA sequencing was done by the dideoxy nucleotide method of Sanger et al. (27). Computer analysis of the sequence data was done with Intelligenetics programs. Subcloned fragments of E51, E33, and E13 were sequenced directly from the plasmid preparations, using a sequencing protocol modified for double-stranded DNA template sequencing. Briefly, 5 μg of CsCl-purified plasmid DNA in 20 μl of water was denatured by addition of 2 μl of 2 M NaOH and allowed to sit for 10 to 30 min at room temperature. The sample was then neutralized by addition of 8 μl of 1 M Tris (pH 4.5) and 3 μl of 3 M sodium acetate and then precipitated by addition of 75 μl of absolute ethanol. After incubation at −70°C for 20 min and centrifugation, the pellet was washed with 70% ethanol, dried, resuspended in 7 μl of water, 2 μl of 5× sequencing buffer, and 1 μl of sequencing primer, and incubated for 30 min at 37°C to allow annealing of the primer. The samples were then sequenced with the Sequenase kit (U.S. Biochemical Corp.) and electrophoresed on 6% acrylamide–8 M urea gels.

Sequences were analyzed with the Intelligenetics package of analysis programs. The hydropathicity profiles were generated by using the PEP program with Hopp and Woods hydropathicity values and a window size of six residues.

Amplification and direct sequencing of proviral integration sites. Primers specific to the U3 region of the LTRs and cellular flanking sequences were made for each clone (E13, E33, and E51). PCR amplifications of proviral-cellular junctions were performed in a total volume of 100 μl containing 300 ng of each primer, 0.5 to 1 μg of target genomic DNA, and 0.2 mM each deoxynucleoside triphosphate (dNTP) in a buffer of 2 mM $MgCl_2$, 100 mM Tris (pH 8.3), 500 mM KCl, and 0.1% gelatin. Cycles were 94°C (1 min), 40 to 48°C (1

min), and 72°C (2 min) for 40 cycles, followed by extension at 72°C for 7 min. PCR products were purified for direct sequencing by centrifugation through a Millipore Ultrafree-100 column (Millipore, Inc.) (1/10 was reserved for analysis on an agarose gel). The entire reaction was lyophilized and resuspended in a 1× buffer for a Taqence sequencing kit (U.S. Biochemical). Half (15 μl) of the sample was used per sequencing reaction. One hundred nanograms of the primer for sequencing (U3 or R region) was end labeled with ^{32}P (5,000 Ci/mmol), and 10 ng of the labeled product was used per sequencing reaction. Reactions were performed according to the directions for the kit, with the following modifications. The PCR product and sequencing primer mixture was boiled for 5 min and cooled to room temperature for 15 min; 8 U of *Taq* polymerase was then added, and 4-μl aliquots were dispensed into each of the dideoxynucleotide termination mixes. The termination reaction mixtures were incubated at 72°C for a total of 20 min, 2 μl of extension mix (7.5 μM each dNTP) was added, and the mixture was incubated for 5 min at 72°C. Samples were then analyzed as described above.

Primers for the E51 analysis were as follows: E51 U3 amplimer/seqmer (ATAAGCTGTGTTTAGTAGTGA), binding to bases 2159 to 2178 of the E51 *Hin*dII-*Eco*RI fragment; E51 flanking amplimer (GTCATGGGTTGATGCAGTATC), binding to bases 2402 to 2382; and E51 R seqmer (GCC ATTTTGCTGCTCATCATAT), binding to bases 2185 to 2206.

Southern blot analysis. The stringency of prehybridization and hybridization was controlled by varying the formamide concentration of the prehybridization and hybridization solutions as previously described (26). Briefly, lowering the formamide concentration by 1% lowers the hybridization complex melting temperature by 1°C; thus, the 50% formamide hybridization solution requires 90% or more sequence identity between probe and target DNA under the conditions described below, while the 30% formamide hybridization solution allows stable probe-target hybrids of 70% or more sequence identity. The standard prehybridization solution consisted of 30 or 50% (vol/vol) formamide (30% formamide for low stringency and 50% for high stringency), 1 M NaCl, 20 mM sodium phosphate (pH 6.8), 0.1% sodium dodecyl sulfate, 5× Denhardt's solution (0.1% polyvinylpyrrolidone 40, 0.1% Ficoll 400, 0.1% bovine serum albumin), and 500 μg of boiled, sheared salmon sperm DNA per ml. Alternatively, the Denhardt's solution could be eliminated and heparin could be added to 500 μg/ml only for high-stringency hybridization with no loss of signal and improved backgrounds when nylon membranes were used (21). Prehybridization, hybridization, and washes were performed as previously described (26). Probes derived from the proviral clones described above were made by random primer labeling of agarose gel-purified DNA fragments (13).

Nucleotide sequence accession numbers. Sequences obtained from the endogenous clones have the following GenBank accession numbers: E51, M95189; E33 LTR, M95190; and E13 LTR, M95191.

RESULTS

Distribution of RSV and EAV-0 *env*-related sequences in the genus *Gallus*. The EAV-0 family of endogenous proviruses is characterized by a complex hybridization pattern on Southern blots, reflecting the approximately 50 proviruses conserved within the domestic chicken genome (11). The distribution and conservation of EAV-0 elements in the genomes

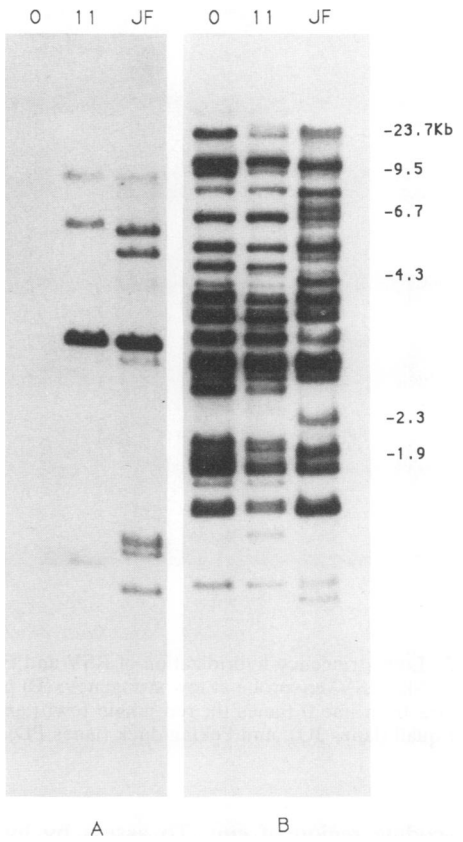


FIG. 1. High-stringency hybridization of RSV and EAV-0 *env* region probes. (A) Southern blot of *Eco*RI-digested genomic DNAs from domestic chicken line 0 (lane 0) and line 11 (lane 11) and red jungle fowl (lane JF) probed at high stringency with the 1.7-kb RSV *env* probe; (B) parallel blot hybridized with EAV-0 *env* region probe at high stringency. Size markers indicated at the right are from *Hin*dIII-digested lambda DNA.

of red jungle fowl, the feral progenitor of domestic chickens, and line 11 and line 0 domestic chickens is in stark contrast to that of RAV-0, which is present at low copy number in most chicken lines and absent in line 0 (1) (Fig. 1). Low-stringency hybridization (allowing hybridization of sequences with 70% identity with the probe) with the EAV-0 and RSV *env* probes was done by using genomic DNAs of Peking duck, ringneck pheasant, peafowl, guinea fowl, Japanese quail, and the *Gallus* species to determine both whether the EAV-0 *env* probe detected all *env*-related elements in the chicken genome and whether the diversity of *env*-related elements detected extends beyond the genus *Gallus* (Fig. 2). The hybridization results indicated that the cloned EAV-0 elements do not account for all ASLV *env*-related sequences in line 0 and that elements at least distantly related to both EAV-0 and RSV *env* genes are present in all avian DNAs tested.

While there are clearly a large number of individual EAV-0 proviruses in the chicken genome, our analysis of the RNA from primary fibroblasts of 12-day-old chicken embryos indicated that the major EAV-0 transcript was shorter than would be predicted by collinearity with RAV-0 (2), suggesting that this major expressed EAV-0 element is defective. Preliminary mapping of independent EAV-0 clones indicated that a major portion of the truncation was in

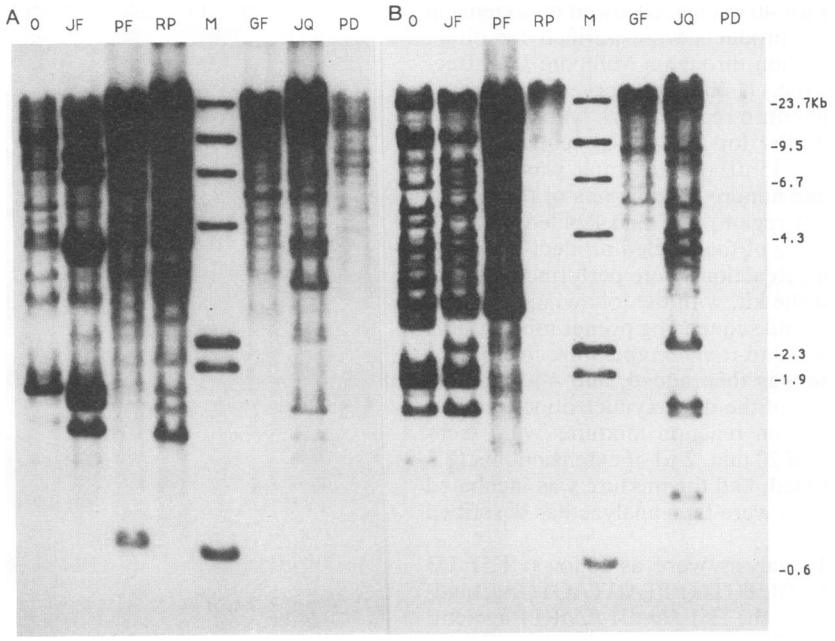


FIG. 2. Low-stringency hybridization of RSV and EAV-0 *env* region probes. (A) Southern blot of *Eco*RI-digested genomic DNA probed with the 1.7-kb RSV *env* probe at low stringency; (B) parallel Southern blot probed with the EAV-0 *env* probe at low stringency. The DNA samples are from line 0 (lanes 0), red jungle fowl (lanes JF), peafowl (lanes PF), ringneck pheasant (lanes RP), guinea fowl (lanes GF), Japanese quail (lanes JQ), and Peking duck (lanes PD). The DNA markers (lanes M) are *Hin*dIII-digested lambda DNA.

the SU-coding region of *env*. To assess by hybridization whether this truncation encompassing the SU (gp85)-coding domain was the predominant deletion in EAV-0 elements, Southern blots of avian genomic DNA were hybridized with the 5′ (SU-coding) half of RSV *env* at low stringency (Fig. 3). Since only two bands were detected in the line 0 genome, this analysis suggests that the majority of EAV-0 proviruses which hybridized to the complete RSV *env* probe are deleted in the SU-coding domain. Note that this pattern of *env* deletion may also exist in related quail endogenous proviruses, since only three fragments in Japanese quail DNA hybridized to the 5′-half RSV *env* probe versus the large number of bands detected in Japanese quail with the complete RSV *env* probe discussed above.

Analysis of the deletion breakpoints in three independent EAV-0 proviral clones. To determine whether EAV-0 elements were deleted in the SU-coding domain or contained SU-coding sequences highly unrelated to RSV SU-coding sequences, we sequenced across the *env* regions in three independently isolated EAV-0 clones. We predicted also that if these deletions were the result of independent integration events or duplication of deleted proviruses, independently generated deletions would be expected to have distinct deletion breakpoints rather than duplication of identical breakpoints. Alignment of the sequences of the deletion breakpoints from the three clones with the sequence of RSV showed that each of the EAV-0 proviruses analyzed was deleted independently (Fig. 4). While duplications of proviruses may also be present in the genome, this result indicates that there are multiple independent germinal integrations of EAV-0 proviruses within the *Gallus* genome and, consistent with the lack of RSV 5′ *env* hybridization described above, that most if not all EAV-0 elements have large deletions in *env*.

Screening for complete EAV-0 *env* genes. To determine

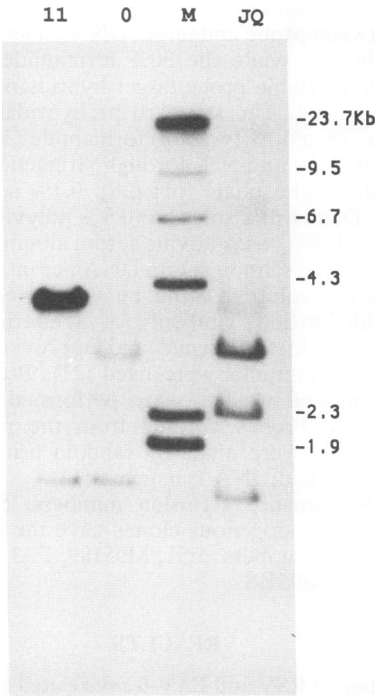


FIG. 3. Southern blot of *Eco*RI-digested genomic DNA samples probed with the RSV 5′ *env* probe at low stringency. The RSV 5′ *env* probe encompasses SU-coding sequences only. The DNAs are line 11 (lane 11), line 0 (lane 0), and Japanese quail (lane JQ). Lane M is as in Fig. 2.

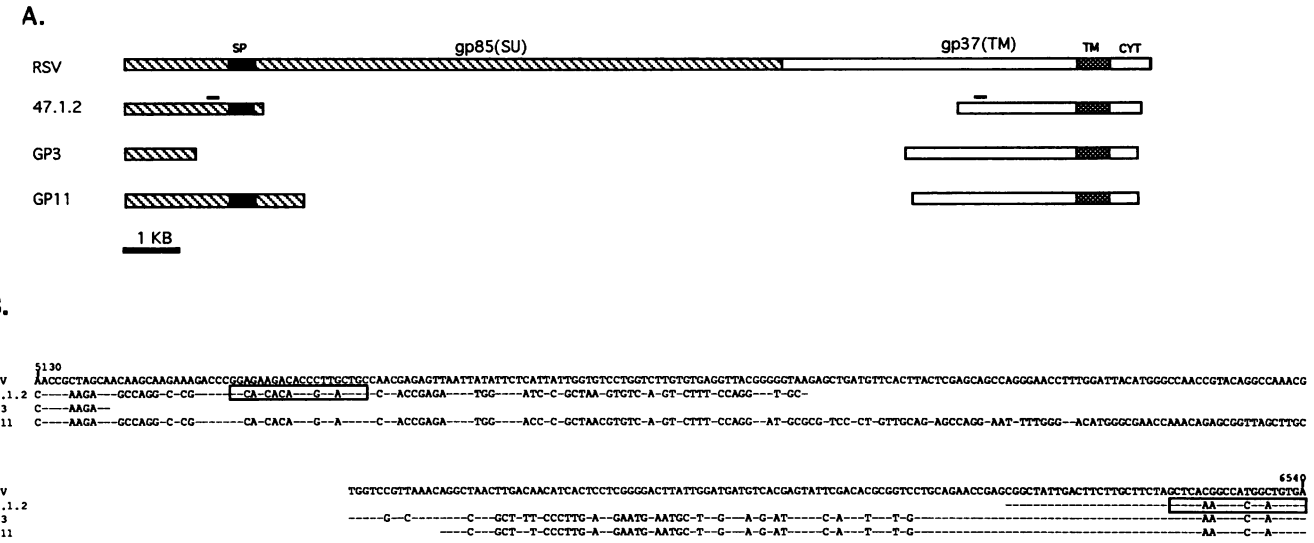


FIG. 4. Deletion breakpoints in EAV-0 proviral *env* genes. (A) Schematic representation of the *env* region deletions for three EAV-0 clones (47.1.2, GP3, and GP11) relative to RSV *env*. The RSV SU signal peptide (SP) is shown as a black box; the transmembrane region (TM) of GP37 is shown as a stippled box and is followed by the cytoplasmic domain (CYT). The black bars over the 47.1.2 sequence show the positions of the PCR primers indicated in panel B. (B) DNA sequence adjacent to the deletion breakpoints of the clones relative to the RSV sequence. Base positions indicated over the RSV sequence are according to Schwartz et al. (29). Boxes indicate primer sequences.

whether any proviruses with complete *env* genes were present in the line 0 genome, a line 0 genomic DNA lambda library was screened with the RSV 5′ *env* probe and an EAV-0 *env* probe under low-stringency hybridization conditions. Low-stringency hybridization was used for the EAV-0 probe as well because it was unclear from the Southern blots whether the bands detected with the RSV 5′ *env* probe were part of the EAV-0 family of elements. Fifteen independently isolated lambda genomic DNA library clones which hybridized to RSV *env* and/or EAV-0 *env* probes at low stringency were obtained. To facilitate screening for full-length *env* genes, each of the 15 lambda clones was analyzed by PCR. The oligomers used to screen the *env* clones were specific to regions immediately adjacent to the deletion breakpoint in EAV-0 clone 47.1.2 (positions of primers are marked in Fig. 4B). PCR analysis was done under relaxed annealing conditions (37°C); no bands were amplified at higher annealing temperatures. PCR results (not shown) fell into three groups. The first group, consisting of eight clones, showed no band on the ethidium bromide-stained gel of the electrophoresed PCR product, indicating the absence of one or both oligomer target sites; the second group (four clones) produced small products (smaller than 1 kb) which indicated major deletions in their *env* genes; the third group, consisting of three clones (E13, E33, and E51), produced products of greater than 1 kb.

While primer choice clearly would influence the ability to detect a divergent *env* sequence, the primers were intended initially to screen for EAV-0 elements that might have full-length *env* genes and the use of lower-annealing-temperature PCR to reduce the possibility of a false-negative result due to sequence divergence. Positive cross-hybridization at low stringency with the RSV and EAV-0 *env* probes indicted that the pool of clones screened should contain regions of at least 70% sequence identity to the corresponding regions of those viruses.

Southern blot analysis of E33 and E51. Two types of hybridization probes were generated from the clones: subcloned fragments which hybridized to RSV *env* at low stringency, and the PCR product generated from each lambda clone by using the EAV-0 *env* region oligomers (Fig. 5). The *env* region PCR product probes from E51 and E33 were used to allow direct comparison of the distribution of *env* gene sequences related to these elements. Identical or closely related high-stringency hybridization patterns from these two probes would indicate that the hybridization pattern accurately reflects the conservation and distribution of *env* genes from these viruses rather than the presence of an unrelated repetitive element integrated within the putative *env* domain of E51 or E33. Hybridization of the E33 and E51 PCR *env* probes to genomic DNA of *Gallus* species relative to the hybridization of the EAV-0 *env* probe showed that the E33 and E51 probes produced a closely related band pattern distinct from that of the EAV-0 probe (Fig. 6). While the EAV-0 *env* probe contains *pol* gene sequences (Fig. 5), the upstream *Eco*RI site that defines the 5′ end of this probe is conserved in all but five EAV-0 proviruses (which do contain a conserved *Eco*RI site 1.3 kb farther upstream) so that the band pattern detected with this probe accurately reflects the distribution of *env* gene sequences and independent integration events, since the downstream *Eco*RI site of each detected fragment is derived from cellular sequences flanking the provirus. The unique hybridization pattern shared by the two new clones, E51 and E33, coupled with the lack of cross-hybridization at high stringency with EAV-0 *env* indicated that E33 and E51 belong to the same distinct group of endogenous sequences. Sequence analysis (see below) indicated that the *env* gene region of E13 is closely related (80% or more identical) to that of E51 but is highly deleted in a fashion similar to deletion of the EAV-0 elements.

Since the E33/E51 group was shown to be distinct from EAV-0, we wanted to determine whether the observed elements are conserved among individual outbred birds as the EAV-0 elements are (26). Southern blots containing *Eco*RI-digested genomic DNA from 14 outbred *G. varius* birds were hybridized under high-stringency conditions with the EAV-0 *env* probe, the E51 PCR *env* probe, and an E51

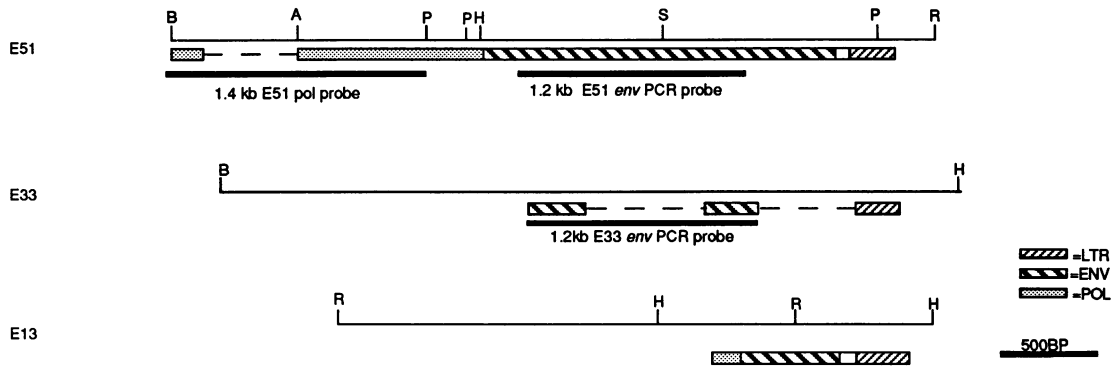


FIG. 5. Restriction maps of proviral clones. The maps of E51, E33, and E13 are derived from the regions of the original lambda genomic DNA clones which hybridized to the RSV and EAV-0 *env* probes. Boxes below each line mark regions identified by sequence analysis. Bars indicate boundaries of probes derived from the clones. Dashed lines indicate regions of proviruses not sequenced. Restriction enzyme sites: *Eco*RI (R), *Bam*HI (B), *Hin*dIII (H), *Pst*I (P), *Xho*I (X), and *Acc*I (A).

pol region probe derived from the lambda clone (Fig. 7). The EAV-0 *env* probe hybridization (Fig. 7A) gave a basic conserved pattern of approximately seven bands in *G. varius* with variation in the inheritance of approximately 12 other bands among the 12 samples. All of the variant EAV-0 bands, with the exception of some in *G. varius* sample 74, are present in more than one sample and in various combinations, and their presence is not associated with the absence of any of the conserved bands, indicating that the variant bands are the products of independent, unlinked proviral integrations which occurred postspeciation.

Hybridization of the same blot with the E51 PCR *env* probe at high stringency produced a unique but conserved band pattern (Fig. 7B). One striking feature of hybridization with this probe is the detection of an approximately 1-kb band in all of the *Gallus* species. The relative intensity of this band suggests that it may represent a common internal fragment of E51-related *env* genes, though neither E51 nor E33 contains a 1-kb *env* region *Eco*RI fragment. The conservation of several other bands across all of the *Gallus* species suggests that this group of endogenous retroviruses is more highly conserved and evolutionarily ancient than the EAV-0 elements; this is the case if the bands represent the proviral-cellular junction fragments, as predicted, since conservation of these bands among the *Gallus* species would indicate inheritance of multiple prespeciation integration events. The conservation of the E33/E51 elements could be seen more clearly when the same blot was hybridized with the E51 1.4-kb *pol* region probe at high stringency, producing a simplified but highly conserved band pattern (Fig. 7C). With this *pol* region probe, a basic pattern of four to five conserved E51-related bands was detectable in all *Gallus* species.

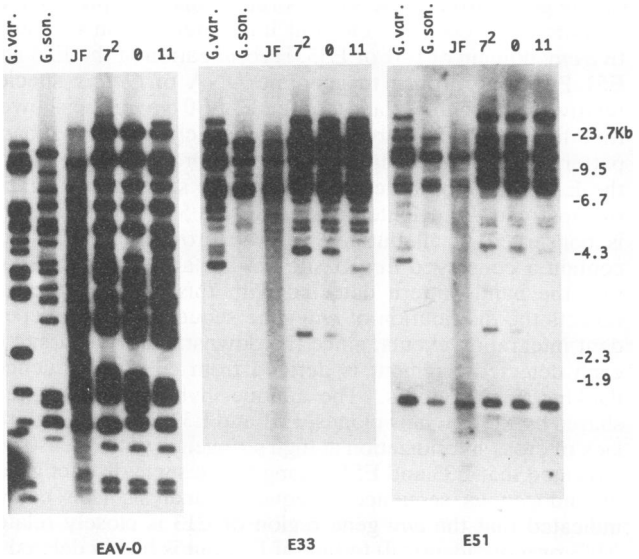


FIG. 6. High-stringency hybridization of E33 and E51 probes to genomic DNAs. A Southern blot of *Eco*RI-digested genomic DNAs was hybridized successively with probes from each of the proviral clones at high stringency. Hybridization with the EAV-0 *env* probe is shown in the left panel for comparison. Probes used were the E33 1.2-kb PCR *env* clone and the 1.2-kb E51 PCR *env* probe. Size markers are as in Fig. 1. DNAs analyzed are from *G. varius* (lanes G.var.) *G. sonerati* (lanes G.son.), red jungle fowl (lanes JF), and domestic chicken lines 7^2, 0, and 11 (lanes 7^2, 0, and 11).

Sequence analysis of E13, E33, and E51. Select regions of the three clones were sequenced to determine their proviral organization and relative similarity to other avian retroviruses. Sequence analysis was performed on the *env*-hybridizing subcloned fragments of each clone; the regions analyzed included the LTRs and all or select portions of the *env* domains.

(i) LTR region analysis. Sequence analysis of the 3′ region of the E13, E33, and E51 clones revealed related LTRs distinct from any previously characterized avian retroviral LTRs in both length and sequence composition (Fig. 8). Comparison of the putative LTR regions with each other and with the EAV-0 LTR allowed delineation of the boundaries of U3, R, and U5. The U3 region of E51 is 242 bp long and is 87% identical to the E13 U3 region. The 5′ region of E51 U3 is also 75% identical to the first 58 bases of EAV-0 U5, which identifies this region as the 5′ terminus of U3. The putative 21-bp E51 R region is 95% identical to that of EAV-0 (one base pair difference from the EAV-0 47.1.2 LTR) but only 52% identical to the collinear region of E13. The 100-bp E51 U5 region is unrelated to the putative U5 region of E13 except for complete identity for six bases at the 3′ terminus. This 3′ terminus of E51 U5 also has 72% identity to the terminal 29 bases of EAV-0 U3. The sequence region of the E33 LTR has approximately 80% identity to that of E51 in the R and U5 regions and 86% identity in the U3 region, excluding three deletions of 5, 12, and 9 bp and

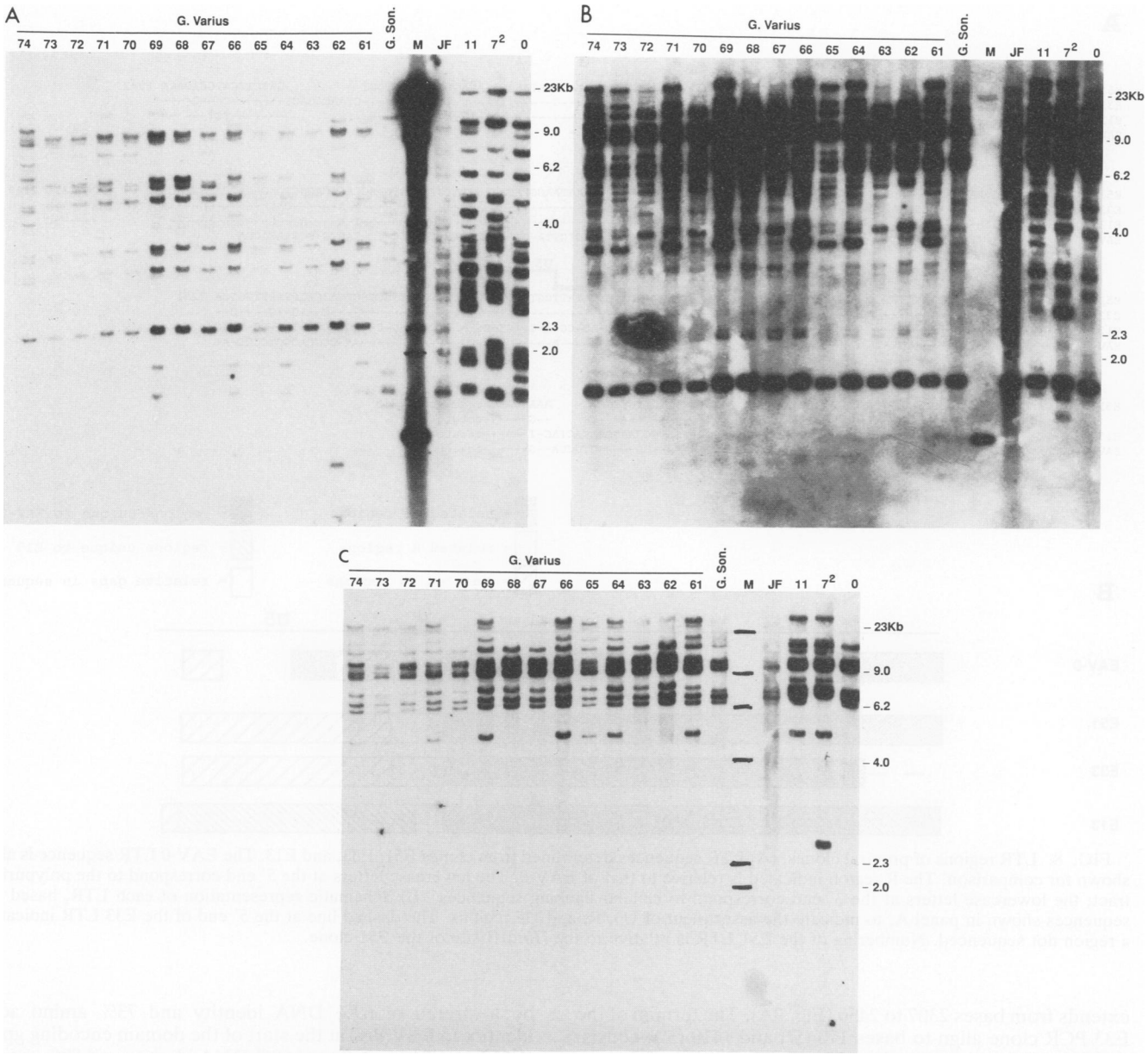


FIG. 7. Conservation of proviral elements among individual birds. A Southern blot of *Eco*RI-digested avian genomic DNAs was probed successively with EAV-0 *env* (A), the E51 PCR *env* probe (B), and the E51 *pol* region probe (C) under high-stringency conditions. *G. varius* samples 61 and 62 are from two unrelated birds from the San Diego Zoo; samples 63 to 74 are from San Antonio Zoo siblings from a cross of unrelated parents (26). The hybridization seen in the marker (M) lane is from internal hybridization controls. In panel C, the markers are indicated with lines drawn in lane M on the basis of prior marker hybridization. Abbreviations are as for Fig. 6.

three insertions of 10, 13, and 9 bp relative to the E51 LTR. This finding indicates that the three clones share related U3 regions distinct from those of all known retroviral groups, R regions most closely related to those of EAV-0 LTRs (except for E13), and divergent U5 regions, with conservation at the site of the terminal inverted repeat. Note that the U5 regions are also longer than those of ASLVs by approximately 20 bases.

(ii) *env* region analysis. Sequence analysis of the *env* gene regions of the proviral clones allowed alignment of the SU- and TM-coding regions relative to each other and to the corresponding sequences of EAV-0 and RSV. The entire 2.5-kb *Hin*dIII-*Eco*RI fragment containing the SU- and TM-coding domains from E51 was sequenced for analysis of potential open reading frames and similarity to other retroviral *env* genes. The termini of the E33 PCR *env* clone also were sequenced to confirm collinearity with E51. The *env* deletion breakpoint region of E13 was sequenced also for comparison with E51 *env*. The putative SU-coding domain of E51 extends from approximately bases 89 to 1147 (counting from the *Hin*dIII site); the TM-coding domain extends from approximately bases 1148 to 1734; the 3′ noncoding region extends from bases 1735 to 1943; the LTR extends from bases 1944 to 2306; and cellular flanking sequence

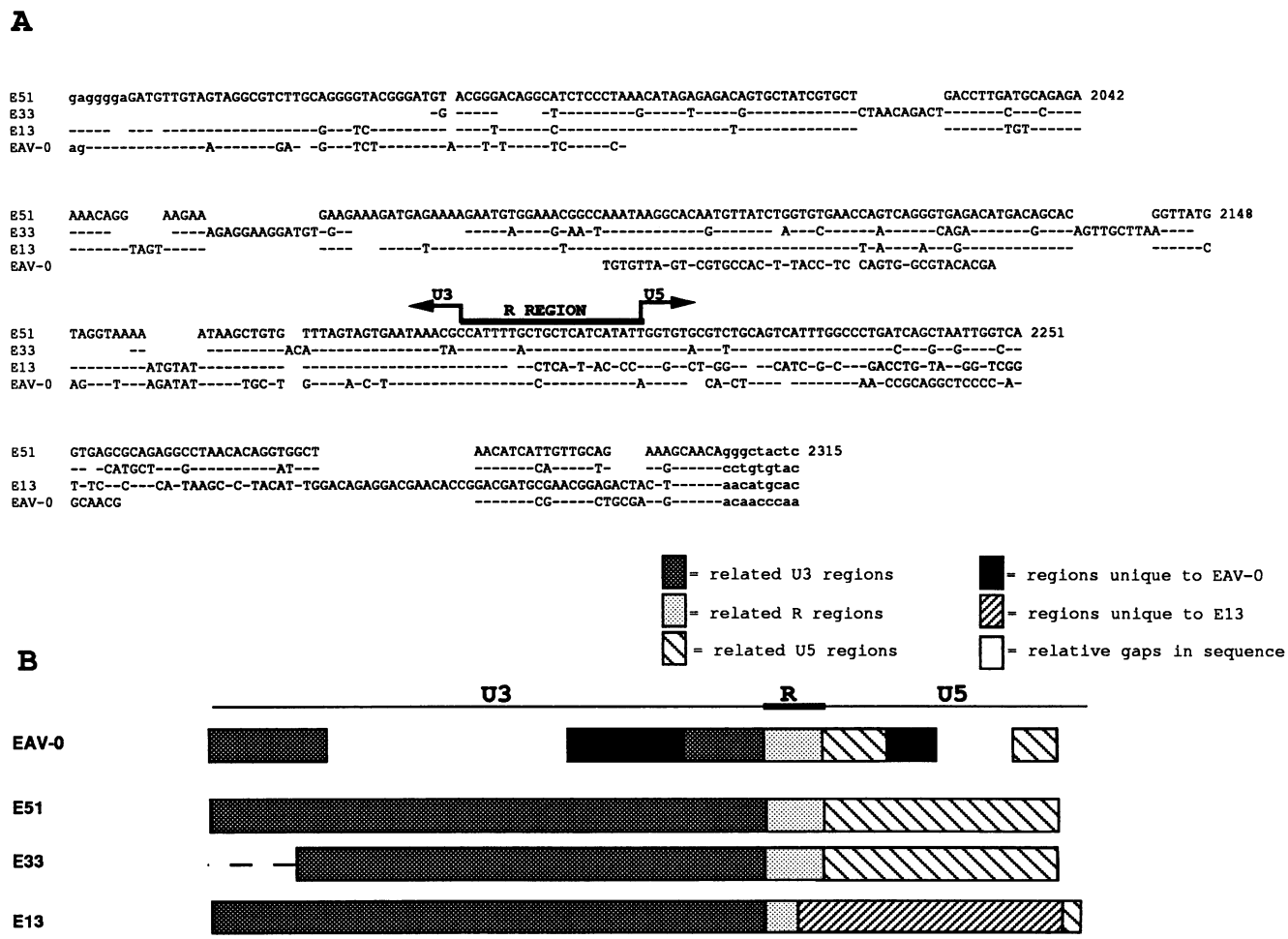


FIG. 8. LTR regions of proviral clones. (A) LTR sequences determined from clones E51, E33, and E13. The EAV-0 LTR sequence is also shown for comparison. The R region indicated is relative to that of EAV-0. The lowercase letters at the 5′ end correspond to the polypurine tract; the lowercase letters at the 3′ end correspond to cellular flanking sequences. (B) Schematic representation of each LTR, based on sequences shown in panel A, to indicate the assortment of U3, R, and U5 regions. The dashed line at the 5′ end of the E33 LTR indicates a region not sequenced. Numbering of the E51 LTR is relative to the *Hin*dIII site of the E51 clone.

extends from bases 2307 to 2450 (Fig. 9A). The termini of the E33 PCR clone align to bases 170 (5′) and 1410 (3′), consistent with the position predicted from the position of the *env* PCR primers (Fig. 9B). Alignment of the E13 *env* gene region indicates a deletion of approximately 1.2 kb extending from base 220, near the start of the SU-coding domain, through base 1463 within the TM-coding domain to the transmembrane region of TM.

Comparison of the DNA and putative amino acid sequences from the E51 SU- and TM-coding domains with sequences of *env* genes of other avian retroviruses allowed approximate delineation of the boundaries of these domains (Fig. 10). The TM-coding domain of E51 is a complete open reading frame with 51% amino acid identity to RSV TM (gp37). The portion of TM encoded by the truncated E13 TM region corresponds to E51 TM amino acids 109 to 197 and has 84% identity to this region of E51 TM. The truncated TM encoded by EAV-0 clone 47.1.2 corresponds to E51 TM amino acids 78 to 197 and has 61% identity to this region of E51 TM. The SU-coding domain of E51 is highly divergent from RSV, with less than 50% identity overall. The boundaries of the putative E51 SU-coding domain can be identified by a stretch of 69% DNA identity and 73% amino acid identity to RSV PrC at the start of the domain encoding gp85 (E51 bases 263 to 362) and 69% DNA identity and 80% amino acid identity close to the end of the gp85-coding domain (E51 bases 973 to 1095).

PCR and direct sequence analysis of integration sites. To determine whether any of the three clones in hand were the result of prespeciation integration, PCR analysis was done on *G. gallus* and *G. varius* genomic DNAs by using U3 and cellular flanking primers for each clone (Fig. 11). Only E51 primers produced amplified product from DNAs other than that of *G. gallus*, and the products from amplification of *G. varius* DNAs were directly sequenced for comparison to the junction sequence of the genomic E51 clone. Direct sequencing of PCR products ensures that differences in sequence are not due to cloning artifacts and that the sequence detected accurately reflects the sequence of the site in the target DNA. Four independently amplified products produced in two different experiments on different days from three different *G. varius* samples were directly sequenced, and all produced the exact same sequence having two base differences and one extra base relative to the genomic E51 clone

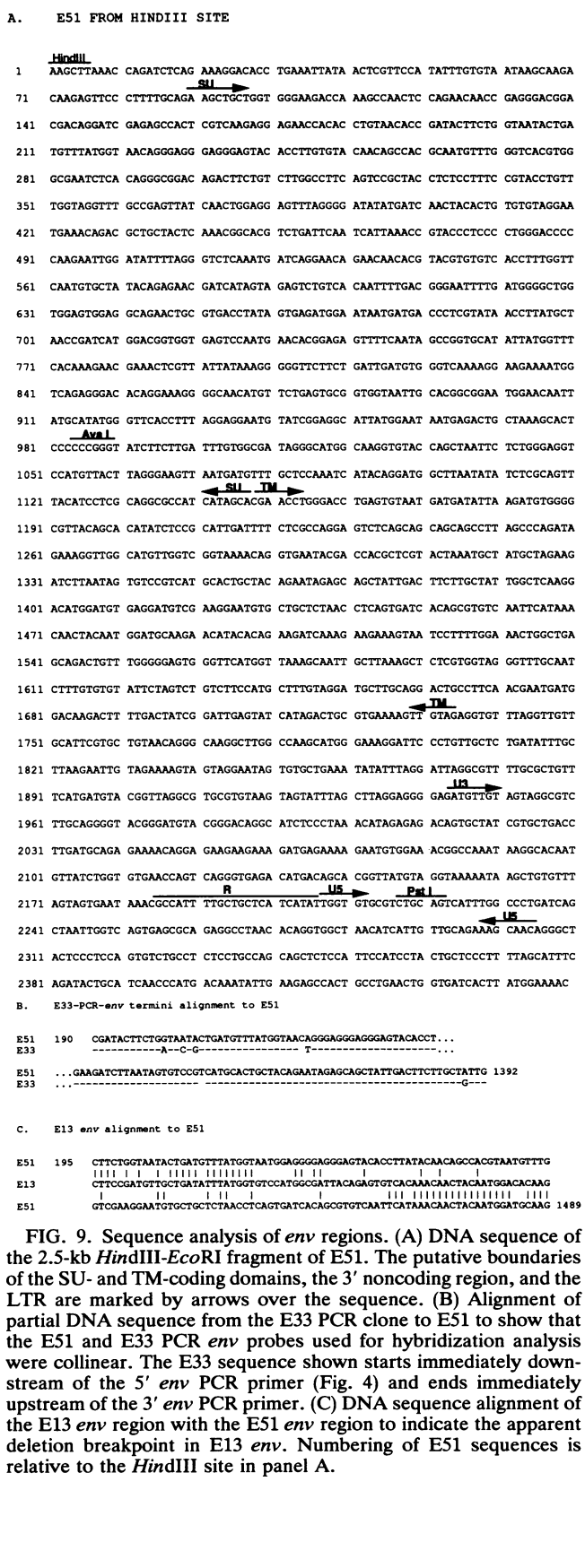


FIG. 9. Sequence analysis of *env* regions. (A) DNA sequence of the 2.5-kb *Hin*dIII-*Eco*RI fragment of E51. The putative boundaries of the SU- and TM-coding domains, the 3′ noncoding region, and the LTR are marked by arrows over the sequence. (B) Alignment of partial DNA sequence from the E33 PCR clone to E51 to show that the E51 and E33 PCR *env* probes used for hybridization analysis were collinear. The E33 sequence shown starts immediately downstream of the 5′ *env* PCR primer (Fig. 4) and ends immediately upstream of the 3′ *env* PCR primer. (C) DNA sequence alignment of the E13 *env* region with the E51 *env* region to indicate the apparent deletion breakpoint in E13 *env*. Numbering of E51 sequences is relative to the *Hin*dIII site in panel A.

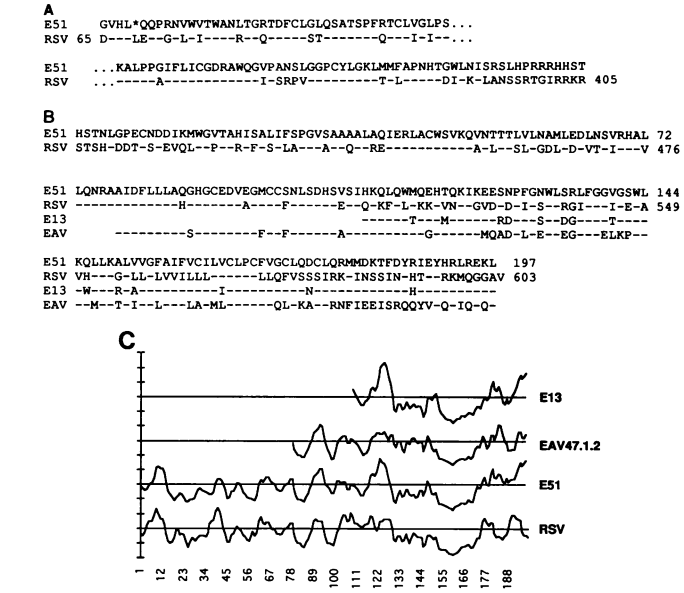
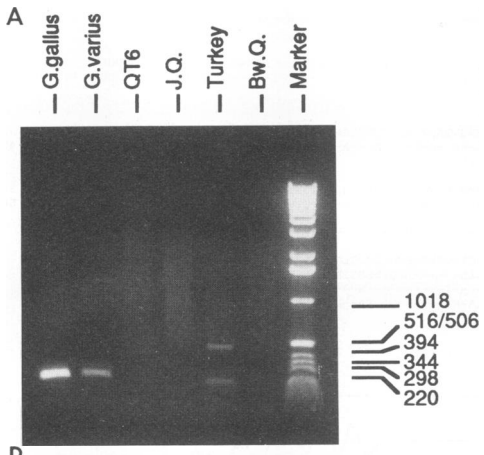


FIG. 10. Amino acid sequence comparison of the TM domain. (A) Alignment of the E51 SU protein sequence at the amino and carboxy termini of the SU domain. The E51 amino acid sequence between these termini had less than 50% identity to RSV SU. (B) Comparison of the E51 TM protein sequence with the RSV TM and the partial TM sequences from E13 and EAV-0 clone 47.1.2. The start points of the latter two sequences correspond to their downstream deletion breakpoints within their *env* genes. Dashes in both panels indicate identity. (C) Plot of the hydropathicity profiles for the amino acid sequences shown in panel B. The *x* axis for each sequence is at neutral (zero) hydropathicity, with hydrophilicity (positive hydropathicity) above each *x* axis and hydrophobicity (negative hydropathicity) below the line. The *y* axis hash marks indicate one-unit increments in hydropathicity values. Numbering of the amino acid positions marked below the graph corresponds to the start of RSV TM. The transmembrane portion of the proteins is centered approximately at amino acid 160.

from *G. gallus*. This result indicates that the E51 provirus integrated prespeciation and that its sequence is highly conserved.

DISCUSSION

Sequence analysis of EAV-0 proviral clones suggested that the bulk of EAV-0 elements present in the domestic chicken genome are deleted in *env*. The independent EAV-0 *env* deletion breakpoints of each clone occurred within 150 bases of each other at both the 5′ and 3′ ends, spanning a region from around the SU signal peptide to the middle of the TM-coding region, resulting in deletions of approximately 1.2 kb in each clone. Each of these cloned proviruses was the result of independent integration events rather than the result of duplication, since each sequenced provirus had a unique deletion breakpoint. The consistent deletion of the SU-coding domain was corroborated by hybridization analysis with the complete and RSV 5′ *env* probes, which showed that there are few DNA fragments related to the SU-coding region of RSV present in the chicken genome relative to the number of fragments detectable with the full-length RSV *env* probe. Such a dramatic absence of SU-related sequences suggested strong selection against functional endogenous *env* genes in the chicken genome. Systematic analysis of a line 0 chicken genomic DNA library



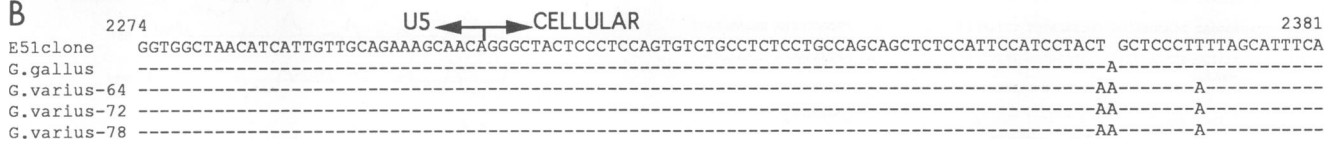


FIG. 11. Analysis of the E51 integration site in *G. varius*. (A) Ethidium bromide-stained agarose gel of PCR products from amplification of target DNAs with the E51 U3 and flanking primers. Target genomic DNAs are indicated as G.gallus, (*G. gallus*), G.varius, (*G. varius*), QT6 (Japanese quail cell line), J.Q. (Japanese quail embryo), Turkey (turkey), Bw.Q. (bobwhite quail), and Marker (size markers). Sizes are indicated in base pairs. (B) Sequence data obtained from direct sequence analysis of the PCR products from three *G. varius* samples (64 [repeated twice], 72, and 78) compared with the sequence obtained from the E51 genomic clone and direct sequencing of the junction fragment from another *G. gallus* sample.

by hybridization and PCR analysis for potentially full-length *env* genes yielded three clones (E13, E33, and E51) with unique sequences. Hybridization analysis indicated that E51 and E33 are closely related to each other and distinct from RSV/RAV and EAV-0. Clone E13 was shown by sequence analysis to have a major deletion in *env* equivalent to that seen in EAV-0 clones. This finding indicated that the original PCR product produced from E13 with the *env* primers may have resulted from the use of an alternative primer binding site as a result of the low annealing temperature (37°C) used.

High-stringency Southern blot hybridization patterns detected with *env* region probes from clones E33 and E51 indicate that many related proviral elements are well conserved among members of different *Gallus* species. Since the probes comprise the same region from two independent proviruses, the hybridization patterns accurately reflect the distribution of *env* gene sequences related to these elements. The E51 and E33 elements also produce highly conserved band patterns across *Gallus* species boundaries, indicating that these elements may be more evolutionarily ancient than the EAV-0 elements. The conserved E51/E33 band pattern comprised four to five bands which were conserved even in *G. sonerati*, suggesting that at least one, and probably several, proviruses entered the germ line prespeciation. Accumulation of many additional proviral elements postspeciation in all of the *Gallus* species indicates that these new elements, as well as EAV-0 elements, were still actively infecting the germ line postspeciation. This hybridization analysis does not unequivocally demonstrate the presence of E51-related proviruses prespeciation, however; therefore, PCR analysis was done to determine whether any of the three clones used in this study integrated prespeciation. Direct analysis of the integration sites of E13, E33, and E51 by PCR showed that E51 entered the germ line prior to divergence of *G. gallus* from *G. varius*. This result indicates that this group of elements predates *Gallus* speciation, as we have previously shown for the EAV-0 elements (26).

One of the most striking features of these new endogenous retroviral elements is the structure of their LTRs. The E13 LTR is characterized by an ambiguous R-region boundary and a uniquely long U5 region. The E51 and E33 LTRs share U3 regions closely related to that of E13 but R regions nearly identical to that of EAV-0. The U5 regions of E51 and E33 are also divergent from that of E13 and are defined by termini related to those of the EAV-0 U5. These distinct combinations of LTR regions suggest that recombination between several distinct avian retroviruses has contributed to the evolution of their LTRs.

Sequence analysis of the E51 *env* region revealed a full-length but nonfunctional *env*-like gene. Comparison of the TM-coding domain of E51 with that of RSV clearly identifies the E51 protein as a TM-like protein. The boundaries of the SU-coding domain are also well defined, but within the core of the SU domain there is little sequence similarity to other retroviral SU proteins or coding sequences. Retroviruses are variously organized according to host range and distinguishing characteristics; this translates into unique sequences in the variable regions of SU and distinct LTRs. Among avian retroviruses, subgroups of ASLVs differ in the SU variable domains, reflecting differences in host tropism, while groups (RAV and RSV) differ additionally in the length and activity of their U3 regions. The proviral elements described here present a mixture of sequence divergence within the coding domains of the elements and the sharing of specific functional sequences within the LTRs, consistent with these elements comprising a distinct group of endogenous retroviruses.

Another unusual characteristic of the LTRs of these endogenous provirus families is their variation in the inverted repeat sequences that define the ends of retroviral LTRs. These inverted repeats form the target for viral IN during proviral integration; IN cleavage results in the loss of a terminal dinucleotide from each end of the provirus, leaving the signature 5′-TG-CA-3′ at the ends of the integrated provirus (6, 8, 12, 24). In all other avian retroviruses, the terminal dinucleotides targeted for cleavage are U3-AA-TT-U5; in the EAV-0 elements, these nucleotides are U3-GA-AG-U5. The E33, E51, and E13 elements also have a GA target at the predicted 5′ end of the provirus, though we have not yet determined the 3′ target from sequencing of a 5′ LTR. The conservation of this distinct target among these

more ancient avian retroviruses suggests that the endogenous IN protein has a distinct target recognition site.

The E13 element is deleted in *env* in nearly the same position as is the EAV-0 *env* deletions, which suggests some strong predilection for deletion of this region, though the presence of the full-length E51 *env* gene indicates that intact *env* genes of this group can be maintained. The broad distribution of the E51-related elements among the *Gallus* species is analogous to that seen for the EAV-0 elements, indicating that there are at least two distantly related groups of endogenous retroviruses highly conserved and broadly distributed within this genus, much in the fashion of the mouse endogenous retroviruses. We currently are determining the extent of expression of these new elements and the functionality of their unique LTRs to elucidate more fully the potential functions of this group.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant CA 18303.

We thank Helen Purvey for expert technical assistance, particularly in mapping and sequencing of the E51 clone.

REFERENCES

1. **Astrin, S. M., E. G. Buss, and W. S. Hayward.** 1979. Endogenous viral genes are nonessential in the chicken. Nature (London) **281**:339–341.
2. **Boyce-Jacino, M. T., R. Resnick, and A. J. Faras.** 1989. Structural and functional characterization of the unusually short long terminal repeats and their adjacent regions of a novel endogenous avian retrovirus. Virology **173**:157–166.
3. **Callahan, R., W. Dronan, D. Gallahan, and L. D'Hoostelaere.** 1982. Novel class of mouse mammary tumor virus-related DNA sequences found in all species of Mus, including mice lacking the virus proviral genome. Proc. Natl. Acad. Sci. USA **79**:4113–4117.
4. **Callahan, R., D. Gallahan, L. D'Hoostelaere, and M. Potter.** 1986. Endogenous MMTV proviral genomes in feral *Mus musculus domesticus*. Curr. Top. Microbiol. Immunol. **127**:362–370.
5. **Cobrinik, D., A. Aiyar, Z. Ge, M. Katzman, H. Huang, and J. Leis.** 1991. Overlapping retrovirus U5 sequence elements are required for efficient integration and initiation of reverse transcription. J. Virol. **65**:3864–3872.
6. **Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis.** 1987. Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. J. Virol. **61**:1999–2008.
7. **Cohen, J. C., and H. E. Varmus.** 1979. Endogenous mammary tumor virus DNA varies among wild mice and segregates during inbreeding. Nature (London) **278**:418–423.
8. **Craigie, R., T. Fujiwara, and F. Bushman.** 1990. The IN protein of molony murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell **62**:829–837.
9. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18s rDNA. Plasmid **13**:31–40.
10. **Dunwiddie, C. T., and A. J. Faras.** 1985. Presence of retrovirus reverse transcriptase-related gene sequences in avian cells lacking endogenous avian leukosis viruses. Proc. Natl. Acad. Sci. USA **82**:5097–5101.
11. **Dunwiddie, C. T., R. Resnick, M. Boyce-Jacino, J. N. Alegre, and A. J. Faras.** 1986. Molecular cloning and characterization of *gag*-, *pol*-, and *env*-related gene sequences in avian cells lacking endogenous avian leukosis viruses. J. Virol. **59**:669–675.
12. **Duyk, G., M. Longiaru, D. Corbrinik, R. Kowal, P. I. DeHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses. J. Virol. **56**:589–599.
13. **Feinberg, A. B., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. Anal. Biochem. **137**:266–269.
14. **Gudkov, A. V., E. A. Komarova, M. A. Nikiforov, and T. E. Zaitsevskaya.** 1992. ART-CH, a new chicken retroviruslike element. J. Virol. **66**:1726–1736.
15. **Hoggan, M. D., C. E. Buckler, J. F. Sears, W. P. Rowe, and M. A. Martin.** 1983. Organization and stability of endogenous xenotropic murine leukemia virus proviral DNA in mouse genomes. J. Virol. **45**:473–477.
16. **Itin, A., G. Potman, and E. Keshet.** 1983. Nucleotide sequence analysis of the long terminal repeat of murine virus-like DNA (VL30) and its adjacent sequences: resemblance to retrovirus proviruses. Virology **127**:347–384.
17. **Jenkins, N. A., N. G. Copeland, B. A. Taylor, and B. K. Lee.** 1982. Organization, distribution, and stability of endogenous ecotropic murine leukemia virus DNA sequences in chromosomes of *Mus musculus*. J. Virol. **43**:26–36.
18. **Katz, R., C. Omer, J. Weiss, S. Mitsialis, R. Guntaka, and A. J. Faras.** 1982. Restriction endonuclease and nucleotide sequence analyses of molecularly cloned unintegrated avian tumor virus DNA: structure of large terminal repeats in circle junctions. J. Virol. **42**:346–351.
19. **Khan, E., J. P. G. Mack, R. A. Katz, J. Kulkosky, and A. M. Skalka.** 1991. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. Nucleic Acids Res. **19**:851–860.
20. **Kozak, C. A., and R. R. O'Neill.** 1986. Xenotropic and MCF-related retroviral genes in wild mice. Curr. Top. Microbiol. Immunol. **127**:349–353.
21. **Kozak, C. A., and R. R. O'Neill.** 1987. Diverse wild mouse origins of xenotropic, mink cell focus-forming, and two types of ecotropic proviral genes. J. Virol. **61**:3082–3088.
22. **O'Neill, R. R., A. S. Khan, M. D. Hoggan, J. W. Hartley, M. A. Martin, and R. Repaske.** 1986. Specific hybridization probes demonstrate fewer xenotropic than mink cell focus-forming murine leukemia virus-related sequences in DNAs from inbred laboratory mice. J. Virol. **58**:359–366.
23. **Ono, M.** 1990. Molecular biology of type A endogenous retrovirus. Kitasato Arch. Exp. Med. **63**:77–90.
24. **Ono, M., H. Kitasato, H. Ohishi, and Y. Motobayashi-Nakajima.** 1984. Molecular cloning and long terminal repeat sequences of intracisternal A-particle genes in *Mus caroli*. J. Virol. **50**:352–358.
25. **Panganabin, A., and H. Temin.** 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. Cell **36**:673–679.
26. **Resnick, R. M., M. T. Boyce-Jacino, Q. Fu, and A. J. Faras.** 1990. Phylogenetic distribution of the novel avian endogenous provirus family EAV-0. J. Virol. **64**:4640–4653.
27. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
28. **Schmidt, M., T. Wirth, B. Kroger, and I. Horak.** 1985. Structure and genomic organization of a new family of murine retrovirus-related DNA sequences (MuRRS). Nucleic Acids Res. **13**:3461–3470.
29. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
30. **Singh, L., and K. W. Jones.** 1984. The use of heparin as a simple cost-effective means of controlling background in nucleic acid hybridization procedures. Nucleic Acids Res. **12**:5627–5638.