

Characterization of a *Xenopus tropicalis* Endogenous Retrovirus with Developmental and Stress-Dependent Expression^{∇†}

L. Sinzelle, Q. Carradec, E. Paillard, O. J. Bronchain, and N. Pollet*

Institute of Systems and Synthetic Biology, Genopole, CNRS, University of Evry, Evry, France

Received 17 September 2010/Accepted 3 December 2010

We report on the identification and characterization of XTERV1, a full-length endogenous retrovirus (ERV) within the genome of the western clawed frog (*Xenopus tropicalis*). XTERV1 contains all the basic genetic elements common to ERVs, including the classical 5'-long terminal repeat (LTR)-gag-pol-env-3'-LTR architecture, as well as conserved functional motifs inherent to each retroviral protein. Using phylogenetic analysis, we show that XTERV1 is related to the *Epsilonretrovirus* genus. The *X. tropicalis* genome harbors a single full-length copy with intact gag and pol open reading frames that localizes to the centromeric region of chromosome 5. About 10 full-length defective copies of XTERV1 are found interspersed in the genome, and 2 of them could be assigned to chromosomes 1 and 3. We find that XTERV1 genes are zygotically transcribed in a regulated spatiotemporal manner during frog development, including metamorphosis. Moreover, XTERV1 transcription is upregulated under certain cellular stress conditions, including cytotoxic and metabolic stresses. Interestingly, XTERV1 Env is found to be homologous to FR47, a protein upregulated following cold exposure in the freeze-tolerant wood frog (*Rana sylvatica*). In addition, we find that *R. sylvatica* FR47 mRNA originated from a retroviral element. We discuss the potential role(s) of ERVs in physiological processes in vertebrates.

Retroviral elements represent a considerable fraction of vertebrate genomes. They have been extensively studied in humans and mice, for which they make up 8% and 10% of the genome, respectively (38, 60). Two classes of retroviruses have been distinguished: exogenous viruses or infectious forms that are horizontally transmitted and endogenous retroviruses (ERVs) that have become incorporated into the host genome following infections of germ or early embryonic cells. ERVs are inherited vertically by the subsequent generations and evolve as endogenous elements of the host genome. They often become defective over time via accumulation of multiple mutations and large deletions and persist as truncated forms or relics within the genome (8).

Exogenous retroviruses are classified in two subfamilies: the *Orthoretrovirinae* and the *Spumaretrovirinae*. *Alpha-*, *Beta-*, *Gamma-*, *Delta-*, and *Epsilonretrovirus* and *Lentivirus* are six genera of the *Orthoretrovirinae* subfamily, and *Spumavirus* is the single genus from the *Spumaretrovirinae* subfamily (9, 23). ERVs derived from these exogenous retroviruses are present in various vertebrate host genomes, including fishes, amphibians, reptiles, birds, and mammals (27). ERV structures can vary to a large extent across these groups, and this is correlated by differences in ERV content between host genomes (8). Thus, a large body of our knowledge on ERVs comes from the study of their diversity in host genomes. However, there is a significant gap of scientific literature on amphibian ERVs.

Three different endogenous retroviral fragments corresponding to retroviral protease and reverse transcriptase genes from the poison dart frog (*Dendrobates ventrimaculatus*) have been isolated and characterized (58). Herniou and colleagues have shown the presence of ERVs in various vertebrate genomes, including several anurans and urodeles (27). Nevertheless, the first and unique complete sequence of an anuran amphibian ERV, Xen-1, was only recently characterized within the *Xenopus laevis* genome (32). Xen-1 is closely related to ERVs derived from *Epsilonretroviruses*, and its presence in the genomes of several species of the *Xenopus* genus, including *Xenopus tropicalis*, has been confirmed (32).

In all species, the majority of ERVs take the form of transcriptionally silent proviral relics. However, some elements retain a certain degree of transcriptional competence, as shown by active promoter and enhancer elements within their long terminal repeats (LTRs) (2, 59). In some instances, ERV regulatory sequences have been coopted by the host to participate in the expression of neighboring genes (39, 50, 57). In addition, increasing data support the view that expression of active ERVs can be precisely controlled and modulated and that ERV products (RNA transcripts/proteins) can be associated with biological functions in both physiological and pathological processes (see reference 30 for a review).

The most striking illustration of the beneficial contribution of transposable element (TE) sequences such as ERVs concerns the molecular domestication process, by which a TE-derived coding sequence gives rise to a functional host gene. Indeed, some ERV proteins have been domesticated to assume important physiological roles. The best examples are the env-derived proteins syncytin-1 and syncytin-2 in primates and their analogous proteins in mammals, which have been recruited to act in placenta morphogenesis (7, 10, 44). Similarly, the *Fv1* (Friend virus susceptibility 1) restriction gene is

* Corresponding author. Mailing address: Institute of Systems and Synthetic Biology, Genopole, CNRS, Université d'Evry Val d'Essonne, Genavenir 3, Genopole Campus 1, 1 rue Pierre Fontaine, F-91058 Evry, France. Phone: 33 0164982748. Fax: 33 0169361119. E-mail: Nicolas.Pollet@issb.genopole.fr.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

∇ Published ahead of print on 15 December 2010.

thought to have evolved from the *gag* region of an endogenous retrovirus related to the MuERV-L element from mice and to the HERV-L element from humans (4, 6). Thus, much like other TEs, the ERV potential for exaptation is high (14).

Here, we report on our discovery and subsequent analysis of a full-length endogenous provirus, named XTERV1, discovered within the *X. tropicalis* genome. We found that this ERV is related to the *Epsilonretrovirus* genus of the *Retroviridae* family. This element exhibited a typical proviral genomic organization and had the distinctive feature of encoding an envelope protein highly similar to a freeze-response protein of the wood frog (*Rana sylvatica*). We estimated the copy number and interspersions of XTERV1 elements by *in silico* studies and chromosome mapping experiments. Since XTERV1 LTRs contained intact regulatory sequences, we explored the temporal and spatial expression patterns of the retroviral genes during development, including metamorphosis. Moreover, we showed that XTERV1 transcription is enhanced by particular cellular stresses, pointing toward intriguing functional analogies with the *R. sylvatica* FR47 protein.

MATERIALS AND METHODS

Animals and biological materials. *X. tropicalis* embryos were obtained by *in vitro* fertilization or by natural breeding using standard methods (11). Staging was according to Nieuwkoop and Faber (45). Oocytes at stages I to VI were obtained as described previously (49), sorted under a dissecting microscope, and pooled (30 oocytes per stage) before total RNA extraction was performed using Trizol reagent (Invitrogen) and Phase Lock gel heavy (Eppendorf).

Preparation of sperm nucleus from *X. laevis* and *X. tropicalis* was adapted from transgenesis methods (36). Genomic DNA (gDNA) from sperm nuclei was extracted as previously described (55).

Cell culture. The Speedy cell line consists of fibroblast-type cells, derived from a primary culture established from an *X. tropicalis* limb bud (called 91.1.F1; a kind gift from H. Y. Hwang, Sanger Institute). Cells were propagated in L15 medium (Invitrogen Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS; Sigma) and a cocktail of penicillin G (50 U/ml) and streptomycin (50 µg/ml) (Invitrogen Gibco). Cells were cultivated in a 28°C incubator and passaged twice a week. For chromosome sample preparations, cells were treated with colcemid (Invitrogen) at 0.6 µg/ml in culture medium for 5 h and fixed according to published procedures (37). These cells were karyotyped, and 21 chromosomes were constantly observed, including a full set of *X. tropicalis* chromosomes plus an additional chromosome 10 (chr. 10; trisomy 10; unpublished data). gDNA from the Speedy cell line was prepared as described above. The XL2 cell line was a kind gift from L. Richard-Parpaillon (CNRS, Rennes, France). Culture conditions were the same as those used for the Speedy cell line.

Cell and embryo treatments. (i) **Cold shock.** Speedy cells were grown to confluence in 10-cm² culture dishes and then transferred to a refrigerator set to 8°C to 10°C for 5 h. A set of 10-cm² confluent plates was cultivated at 28°C and used as an untreated control. Cells were harvested by trypsinization, pelleted, and resuspended in 500 µl Trizol reagent (Invitrogen) for total RNA extraction. *X. tropicalis* tadpoles (stage NF47-48) were incubated at 4°C for 1 h, and 20 individuals were pooled for RNA extraction.

(ii) **TH treatment.** Speedy cells grown to confluence in 10-cm² culture dishes were transferred in L15 medium, 10% FBS containing 10 nM thyroid hormone (TH; T3,3,5,3'-L-triiodothyronine). Cells were collected after 8 h and resuspended in 500 µl Trizol reagent (Invitrogen). *X. tropicalis* tadpoles (stage NF47-48) were treated with 10 nM TH and incubated for 17 h and 42 h. Twenty embryos were pooled for total RNA extraction. Untreated embryos served as controls.

(iii) **Serum starvation.** Speedy cells were grown in 10-cm² culture dishes at 80% confluence. Growth medium was removed, and cells were washed twice with phosphate-buffered saline. Serum-depleted medium, which consisted of L15 medium supplemented with 0.2% heat-inactivated FBS, was then added. After 20 to 24 h of treatment, cells were harvested by trypsinization, pelleted, and treated for total RNA extraction.

(iv) **UV-C irradiation.** Speedy cells grown in 10-cm² culture dishes were overlaid with 10 ml of phosphate-buffered saline, placed in a UV cross-linker (BXL-254; Bio-Link), and exposed to 3 mJ/cm² of UV-C irradiation. Phosphate-buff-

ered saline was then replaced with growth medium, and cells were collected at 3 h postirradiation for total RNA extraction.

Real-time reverse transcription-quantitative PCR (RT-qPCR). Total RNA was extracted from pools of 10 to 15 embryos in various developmental stages using Trizol reagent (Invitrogen) and Phase Lock gel heavy (Eppendorf). After DNase treatment, RNA samples were purified with either a MEGAclear purification kit (Ambion) or an RNeasy MinElute kit (Qiagen). RNA integrity was evaluated using an RNA 6000 nanokit on an Agilent 2100 bioanalyzer. One microgram of total RNA was subjected to *in vitro* reverse transcription using a mixture of poly(dT) and random pentadecamer primers (SuperScriptIII; Invitrogen) (46). Products obtained from reverse-transcribed RNAs were monitored using an RNA 6000 picokit on an Agilent 2100 bioanalyzer.

Real-time amplification was conducted in triplicate using the Maxima SYBR green/carboxy-X-rhodamine (ROX) qPCR master mix (Fermentas) on a StepOne apparatus (Applied Biosystems). Biological replicates were performed for the following experiments: serum starvation, UV irradiation, as well as the temporal expression of XTERV1 transcripts. Technical replicates were done for other experiments. The equivalent of 5 ng of total cDNA was subjected to 40 cycles of amplification in two steps: 95°C denaturing for 15 s and 60°C annealing for 1 min. A melting curve analysis was then conducted for 15 s at 60 to 95°C. The various primer sets are listed in Data Set S1 in the supplemental material. We sequenced the amplicons obtained by qPCR using Pol and FR47/Env primers on cDNA samples prepared from two developmental stages (NF48 and NF60-NF62) and on gDNA samples prepared from Speedy cells and an *X. tropicalis* individual. The sequences were identical between them and to the sequence of the XTERV1 full-length copy, which confirms that our qPCR assay is specific to the XTERV1 full-length copy. The normalization of the input cDNA was performed on all samples by monitoring transcript levels of housekeeping genes, such as *X. tropicalis* ornithine decarboxylase 1 (*odc1*) or/and *X. tropicalis* ribosomal protein L8 (*rpl8*) (54). The comparative threshold cycle (C_T) method was used to determine the relative gene abundance (42). The fold change in target gene expression at a given time was calculated for each sample as the ratio of target mRNA normalized to that of the reference mRNA (*odc1* or *rpl8*) and relative to the expression at time zero, using the following formula $2^{-\Delta\Delta C_T}$. Data were expressed as means \pm standard errors of the means (SEMs).

PCR amplification and cloning. (i) **Sequencing of the gap in XTERV1 ORF1/*gag* and genomic assembly of XTERV1 sequence.** The genomic sequence of XTERV1 was incomplete because a gap interrupted its sequence. The sequences bordering this gap on scaffold_387 could be aligned with 92% identity with a 2,117-bp fragment of the scaffold_688 sequence. This gap was most likely due to a misassembly of the end portions of scaffold_387 and of scaffold_688. Primers flanking the lacking sequence in XTERV1 *gag* were used to PCR amplify the gap fragment (primers 5' ORF1 seq1 and 3' ORF1 seq1; see Data Set S1 in the supplemental material). PCR was performed on Speedy cell gDNA (300 ng) diluted in PCR buffer (Fermentas) containing 200 µM each deoxynucleoside triphosphate, 2 mM primers, 1 mM MgCl₂, and 1 U of *Pfu* DNA polymerase (Fermentas) in a 50-µl reaction volume. The cycling procedure was 94°C for 5 min with a hot start and then 35 amplification cycles (94°C for 1 min, 55°C for 1 min, 72°C for 2 min), followed by a final extension step at 72°C for 10 min. The resulting amplicon was gel purified using a QIAquick gel extraction kit (Qiagen), cloned into the pCR2.1-TOPO vector using a TOPOTA cloning kit (Invitrogen), and then sequenced (GATC, Germany). This sequence of 1,232 bp allowed us to fill the gap in the XTERV1 sequence and to assemble an uninterrupted XTERV1 proviral sequence. We assessed the colinearity at a single locus of our assembly of the XTERV1 genomic sequence using PCR on gDNA from Speedy cells or individuals. Using primers 5' QP2 ORF1 and 3' Pol seq1, we amplified a fragment spanning the gap and a 2-kb fragment flanking the 3' end of the gap. For each gDNA template, we obtained and sequenced, as described above, a single amplicon of 3,243 bp. These two sequences were identical to our assembled XTERV1 proviral sequence, confirming its existence at a single locus.

XTERV1 proviral sequences were amplified from gDNA extracted from two unrelated individuals originating from two different populations of *X. tropicalis* (Uyere, Nigeria, and Adiopodoume, Ivory Coast), as well as from the Speedy cell line. The sequences obtained were identical to the XTERV1 reference genomic sequence (GenBank accession no. HM765512), except for two adenines within the *gag* nucleotide sequence (position 2702 in the sequence in Data Set S3 in the supplemental material) and absent from the PCR products obtained from gDNA.

(ii) **Cloning of XTERV1 ORF3/FR47/*env*.** A short fragment (1,305 bp) and a long fragment (1,718 bp) containing the FR47/*env* open reading frame (ORF) were PCR amplified from Speedy cell gDNA using the conditions described above and primers pair 5' FR47 seq3 and 3' FR47 seq1 and primers pair 5' FR47 seq2 and 3' FR47 seq1, respectively. Amplicons were purified using a QIAquick

gel extraction kit (Qiagen) and ligated into the PCRII-TOPO vector using a TOPOTA cloning dual promoter kit (Invitrogen). The resulting plasmids, pCRII-TOPO(FR47/env short) and pCRII-TOPO(FR47/env long), respectively, were validated by sequencing.

(iii) **Amplification of XTERV1 3' junction.** The endogenous nature of XTERV1 was checked by PCRs performed on gDNA extracted from sperm nuclei using the following primers: 3' locus Xterv and 5'QPCR FR47. These primers were designed to amplify a 1,614-bp amplicon corresponding to the 3' junction of XTERV1. The PCR conditions were as described above. The amplicon was gel purified using a QIAquick gel extraction kit (Qiagen) and directly sequenced.

(iv) **Amplification of *Xenopus laevis* XLERV1 pol ORF.** PCR was performed as previously described on gDNA extracted from a *X. laevis* sperm nucleus preparation by using primers 5' XLERV seq1 and 3' XLERV seq1. The 2,635-bp amplified DNA was gel purified and partially sequenced on both extremities.

FISH. The template DNA (500 ng) was labeled by random priming (Decalabel DNA labeling kit; Fermentas) with incorporation of digoxigenin-11-UTP (DIG-11-dUTP) nucleotide (Roche). Purification of the reaction mixture was done using a gel extraction kit (Qiagen). Fluorescence *in situ* hybridization (FISH) coupled with tyramide signal amplification (TSA) was carried out according to published methods (37). In brief, the slide preparation as well as labeled probes were denatured for 5 min at 70°C in hybridization buffer (50% formamide, 2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 300 mM NaCl, 30 mM sodium citrate, pH 7.0) and incubated together overnight (12 to 16 h) at 37°C. Visualization of the hybridized probe was performed using an antidigoxigenin-peroxidase, Fab fragment antibody (Roche). Amplification of the FISH signals was carried out with a TSA-tetramethylrhodamine kit (NEN, Life Science, Boston, MA). Pictures of metaphase spreads after FISH were taken under a fluorescence microscope equipped with an AxioCam MRm camera and processed through AxioVision software (Zeiss, Germany). Labeled chromosomes were identified using their *p/q* arm ratios and relative sizes. The total length of the largest chromosome, chromosome 1, was used as an internal standard (34). Two slides per probe and over ≈30 metaphase spreads per slide were analyzed.

Whole-mount *in situ* hybridization (WISH). *In situ* hybridization studies were carried out according to standard procedures and as previously described (48). Templates for *in situ* probes were obtained by PCR using M13 forward and reverse universal primers on pCRII-TOPO (FR47/env short). Digoxigenin-labeled RNA antisense and sense probes were transcribed with T7 and SP6 RNA polymerases, respectively, using digoxigenin-11-UTP (Roche). Hybridizing probes were detected using an anti-DIG alkaline phosphatase (AP)-conjugated antibody (Roche). Following hybridization, the signal was revealed using the BM purple AP substrate (Roche). For whole-mount imaging, embryos were bleached in 1% H₂O₂-5% formamide-0.5× SSC and then progressively transferred to methanol and viewed in Murray's clearing solution (benzyl benzoate and benzyl alcohol, 2:1 [11]).

Sequence analyses and phylogeny. ORF determination was carried out using ORFfinder from NCBI (www.ncbi.nlm.nih.gov/ORFfinder). Meaningful alignment between the *R. sylvatica* FR47 ORF and *X. tropicalis* FR47-like ORF required the introduction of a frameshift within the *X. tropicalis* FR47-like sequence (Fig. 1A, position 669) and the change of the initiation codon within the *R. sylvatica* FR47 sequence (the first methionine, at position 260 in Fig. 1A) defined elsewhere (43). The identification of conserved functional motifs was made using LTR Finder (61). Reverse transcriptase domains (Prosite PS50878) from representative members of the seven retroviral genera were extracted and then aligned using the MUSCLE program (21). ProtTest was used to select the RtREV+I+G amino acid substitution model before phylogenetic reconstruction using the PhyML program (25). Bootstrap values were determined from 1,000 replicates. LTR sequences were recovered and aligned using the NCBI Blastn and BlastAlign programs (3). For the insertion time estimate, LTR sequences that contained 45% of the gaps in the final alignment were excluded for further analysis. All positions containing gaps and missing data were eliminated, resulting in a final data set with 151 positions from 79 sequences. The T92+G nucleotide substitution model was selected from a maximum likelihood fit analysis of 24 models, and the value of the gamma shape parameter (G) was estimated before estimating the number of base substitutions per site (*d*, which is equal to 0.0848 ± 0.0164). For the LTR phylogeny, the model used was T92+G, in which G was equal to 0.5553, and the 22 sequences provided 96 positions. Analyses were made using the MEGA (version 5) program (56).

Nucleotide sequence accession number. The nucleotide sequence of XTERV1 has been submitted to the EMBL, GenBank, and DDBJ sequence databases under accession number HM765512.

RESULTS

Identification of retroviral envelope protein homologs in *R. sylvatica* and in *Xenopus*. We previously identified a transcript highly expressed during *X. laevis* metamorphosis that exhibits significant similarity to the freeze-responsive FR47 mRNA initially discovered in *R. sylvatica* (22, 43). FR47 encodes a 390-amino-acid (aa) protein specifically expressed in the liver of this freeze-tolerant wood frog, suggesting that it may be involved in freeze survival (43). In *X. laevis*, the FR47-like transcript is predominantly expressed in metamorphic limbs. However, transcripts are also detected during embryogenesis (starting at the gastrula stage) and in most adult tissues sampled, the exceptions being the ovary, testis, and lung (22).

We analyzed the corresponding loci available in the *X. tropicalis* genome sequence to further characterize this FR47-like gene (26). The scaffold_387 was predicted to contain an ORF (at positions 1123386 to 125395) potentially encoding a protein with significant similarities to FR47 from *R. sylvatica* (cDNA, GenBank accession no. AY100690). The predicted 742-aa *X. tropicalis* FR47-like protein aligned over its whole length with the 649-aa putative FR47 protein from *R. sylvatica* (Fig. 1A), and the two sequences share 36% similarity. Further investigations revealed that these two proteins displayed significant similarities with other known retroviral envelope proteins, including the ORF3 of the chicken (*Gallus gallus*) ERV OVEX1, its zebra finch (*Taeniopygia guttata*) homolog (13), and a portion of *env* from the grouse (*Bonasa umbellus*) ERV (19), as well as with genomic and cDNA sequences from various vertebrate species (Fig. 1B; see Data Set S2 in the supplemental material). Moreover, sequence analysis of *X. tropicalis* FR47-like and *R. sylvatica* FR47 proteins revealed the presence of conserved motifs typically described in envelope proteins, including a hydrophobic signal peptide at the N terminus, a transmembrane domain at the C terminus, multiple putative glycosylation sites, as well as a furin cleavage site (consensus R/K-X-R/K-R), which separates the surface region (SU) and the transmembrane domain (TM) subunits (5).

Altogether, we showed that the conceptually translated product of the *X. tropicalis* FR47-like sequence corresponds to a bona fide 742-aa retroviral envelope protein (referred to as the FR47/Env ORF). *R. sylvatica* FR47 and *X. tropicalis* FR47/Env belong to a family of envelope-like proteins, which are widespread in vertebrates, from amphibians to mammals.

In 2003, McNally et al. positioned the coding region of FR47 at the C-terminal region of the translated product predicted from the 3,678-bp sequence of *R. sylvatica* cDNA (43). However, we found that the 5' end of this cDNA resembles a retroviral polyprotein. Indeed, the N-terminal 401 aa of the putative protein encoded from this cDNA sequence had 28% identity to the Gag-Pro-Pol polyprotein of the walleye dermal sarcoma virus (WDSV). Therefore, we conclude that the FR47 mRNA described by McNally et al. (43) is derived from a retroviral element within the *R. sylvatica* genome.

Characterization of an *X. tropicalis* endogenous retrovirus, XTERV1. We analyzed the genomic sequence containing the FR47/Env ORF and found 485-bp LTRs on both sides (Fig. 2; see Data Set S3 in the supplemental material). The two LTRs are 99.8% identical to each other, with a single substitution

A.

FR470_RANSY 1 -----M N F L D I T V L E C L L G C I Y A F T P G I D V T T H E N N ----- 33
 FR47L_XENTR 1 M L T R L Q K M T S S H L I T S P O D V R T R M P R M T R R M T R E K L N Q L P V M N R I S T K Y 50

FR470_RANSY 34 -----T D L T W G F G P K D G T R S Y S V G S I V V S G H E S P F Y V K D 68
 FR47L_XENTR 53 A L W C S L L T F L L F I P L W I L L N C V C Y P S Q N S P L G S L C Y N A M R N S N N T E H V R K 104

FR470_RANSY 69 N D Y P R S P V V F V K A R P P A L V N -----I S F H W R -----E N L P W P Y N Y T P H K 108
 FR47L_XENTR 105 S L S T M L S S T F M V H D G T D N F S P A N I M Q W G W R N A T I A R Q D N F T C P A N Q C I V 156

FR470_RANSY 109 C H K L G P L E T W K E V F D K T Q A V L R T P H S E E Y K K T K G T G I R P V S G A E P F N I T G P 160
 FR47L_XENTR 157 L N M T T S S S I L K S H N H I G Y L A I P V Q R Y C N F S I I S N A T A F Q S T L L M P S Q E Y D K P 208

FR470_RANSY 161 I V T N Q S A T Y I C A S W F F V K S N N T A G G Y I Q W E H Y G F S I K -----I E D T 201
 FR47L_XENTR 209 F L V N V Y H D P N S F N W T Q N T N Y S L P A G Y T L L F T N G T P I T N R T W F M I G I V C I I S T 260

FR470_RANSY 202 P P L V F K S S V L T P V P K S C L N T S I S Q Q K T K V K F N V T F P K L P D C -----R Y K R E W 248
 FR47L_XENTR 216 S P T A H N - Y S L I P T P D T C V N V T I T I N K A V K V I F K I T T P I P V C S T S K S R H K R A W 311

FR470_RANSY 249 Y D T V L G A T G T G M G V M S G I Q M E M I Q N K W H A G H I A D G L I I T S Q W L P T T F E T Q 300
 FR47L_XENTR 312 Y D T L L G G A G T A M G V L N G V D M E V L R N K L S T A A S H L K T A F H V T A Q W A P T V V Q S Q 363

FR470_RANSY 301 I S Q V L T K Y L N A M V N Y T I A S L E L W R R D Q K F I N L T Q C A F S T L S Y H I Q V N R A R 352
 FR47L_XENTR 364 L M Q L K D Q E F V K Y L G D N T F A V L N L T A G S - - - V S W T L G L R Y V V L S Q R S E L E 412

FR470_RANSY 353 E E L Q S G N Y H T W L Y K F K G L D L E N T W F E V P G R E I E C E E Q W - - - - - C A G R F N 396
 FR47L_XENTR 413 N M L L L P K Y S K L A K L F S W N S R D - W H L D K I N T F C Q E I H N G S I S F F T C Y L S I T 462

FR470_RANSY 397 I Y M V E G I E V M C K L I V M P L L V G E D P P E - - - F W P Y E I Y G H Y A D T Q D R - - - - - 438
 FR47L_XENTR 463 L Y K A D N V S L A C K F T V L P P H Y N G L T S N L T W W P T F L G N Y V F P K T N R T F L E H 514

FR470_RANSY 439 -----T H D L C L L K H Q S N I C K F Q R M P V H I G N R F M E I G P Q H I C L 475
 FR47L_XENTR 515 C I S T N N G L A C R H S S P V Y E P C L H K D L N I C H W D V Y P A D Y S - I M L E I G P Q Y V C F 565

FR470_RANSY 476 I T N D N D T M R S L N R T A P F S G C I M N V K V F K W L N D T F I F E S D A D K T F N R E W A V D N 527
 FR47L_XENTR 566 A T N F P A G I H A V D S S L P F S G C L T N V T L L H W Y N T Y Y F S P I P T V H R S F V Y F P P T 617

FR470_RANSY 528 L T D T T P F V I S L E P L L Q V L K E S I L R K Y I E T H E Y L R N N L L S A I D K G K L I H 579
 FR47L_XENTR 618 L P V G E Y Q I P N A V Y Q S L I I H N T N M I A A Y R K - H D T Q L H D T A I K T V A T K N K L L S 668

FR470_RANSY 580 S S Q I K E E T H H W D V F S G W S P A T K T F S M I F S P I L I L I L G L A I V T V I N C C I Y 631
 FR47L_XENTR 669 M - - - - D E L Q H H W D V F T G W S P T G S A I M Q S F E Y P L A V I A V L V L L V L F N L Y S 716

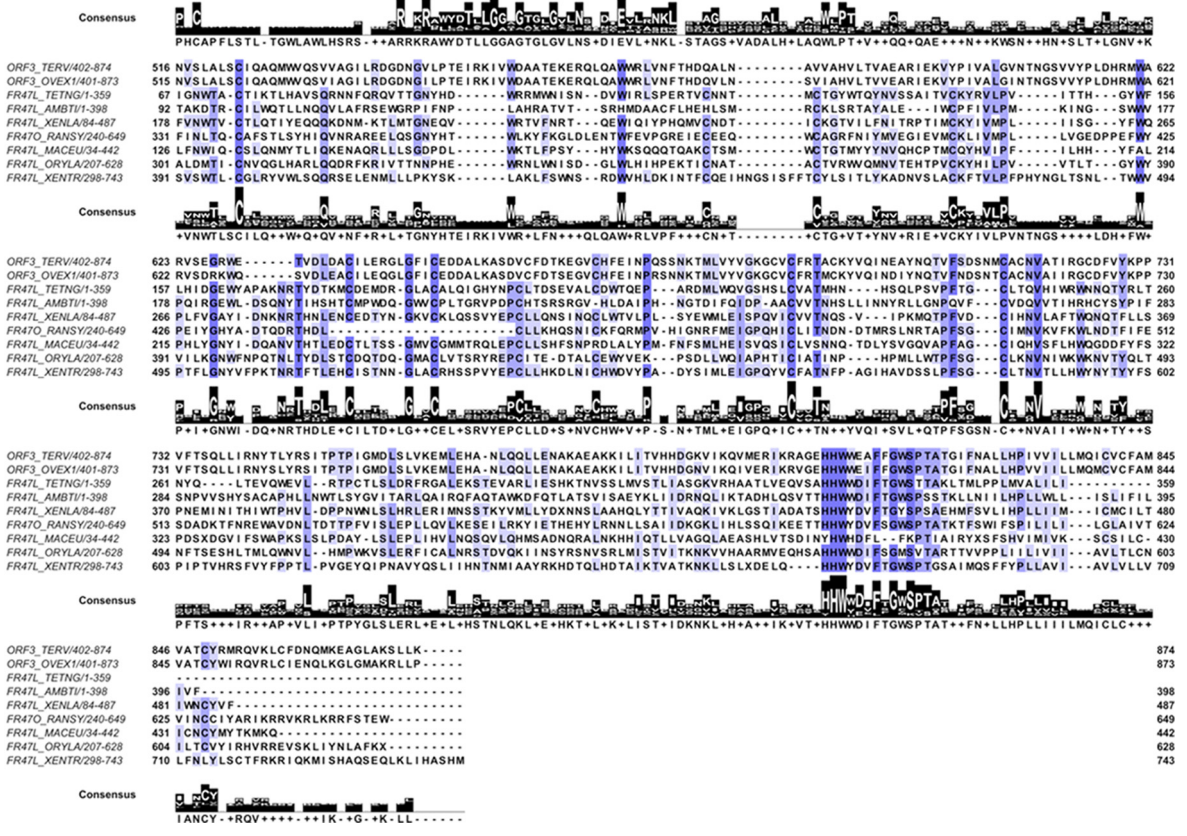
FR470_RANSY 632 A R I K R R V K R L K R R F S T E W ----- 649
 FR47L_XENTR 717 C T F R K R I Q K M I S H A Q S E Q L K I H A S H M ----- 743

SU
TM

B.

ORF3_TERV1402-874
 ORF3_OVE1401-873
 FR47L_TETNG1-359
 FR47L_AMBT11-398
 FR47L_XENLA/84-487
 FR470_RANSY/240-649
 FR47L_MACEU/34-442
 FR47L_ORYLA/207-628
 FR47L_XENTR/298-743

402 P H C A P F L S T L T H G W L A W L H S R S L O G T A R R D L L A T A L G G A A G L G V L N S M D A E V L A N K L - E T V T L G V Q D L K P L N S S L S L G M G Q W L V S E V L P T W E Q I S E K D H Q V L L R A L G M E Q S 515
 401 P H C A P F L S T L Y T G W L A W L H S R S M Q E A T E R D L L A T A L G G O G A G L G V L N S M A E V L A N K L - E A V T S Q V Q L N P L N S S L T S L G M G Q W L V S E V L P T W E H I S E K D H V L L Q A L C I E G N 514
 1 - - - - - T K L V A R G L D S L I A G A A G V T G A L N A F D I E T V R N K L - A T M S D A S G A I K L L A Q W P L I S P L R T N A T S I G N H W Q N F S Y E D I F S A I G Q V N K 91
 84 R S C - - - - - H N A R K R A W D T L L G G Y T L A E T L N S F D I E T L A N R M - H K A G G L T N D A V V W A Q A K M P T I W A P L Q Q V E I D K N I N M M E L E V N F S L N F D N T M T K 370
 240 P C - - - - - R Y K R E W D T V L G A T G T G M G V M S G I Q M E M I Q N K W - H K A G H I A D G L I I T S Q W L P T T F E T Q I S O V Q L T K Y L N A M V N Y T I A S L E L W R R D Q 133
 240 P N C - - - - - A R I K R A W D T L L G G T G L G I V N S I A L E T L A S R L - R S A G D V A Q A L T V N A K W P T M I L P H Q T L E Y O G Q L Q V F N D S L T A I T F N A V T A 125
 207 P A C - - - - - H G V R K R A W D T L L G G A T I L G L S N T A N T E V T R S M L - S S A G D A S Q A L H S V G A W P T V L T Q Q A A K V F F A Q F G W T V T A N E A A T V L K N V T K 300
 298 P V C - - - - - S T S K S R K R A W D T L L G G A T A M G V L N G V D M E V L R N K L - S T A A S H L K T A F H V T A Q W A P T V V Q S Q L M Q L K D Q E F V K - - - - - Y L G D N T F A V L N L T A G 390



ORF3_TERV1402-874
 ORF3_OVE1401-873
 FR47L_TETNG1-359
 FR47L_AMBT11-398
 FR47L_XENLA/84-487
 FR470_RANSY/240-649
 FR47L_MACEU/34-442
 FR47L_ORYLA/207-628
 FR47L_XENTR/298-743

874
 873
 398
 487
 649
 442
 628
 743

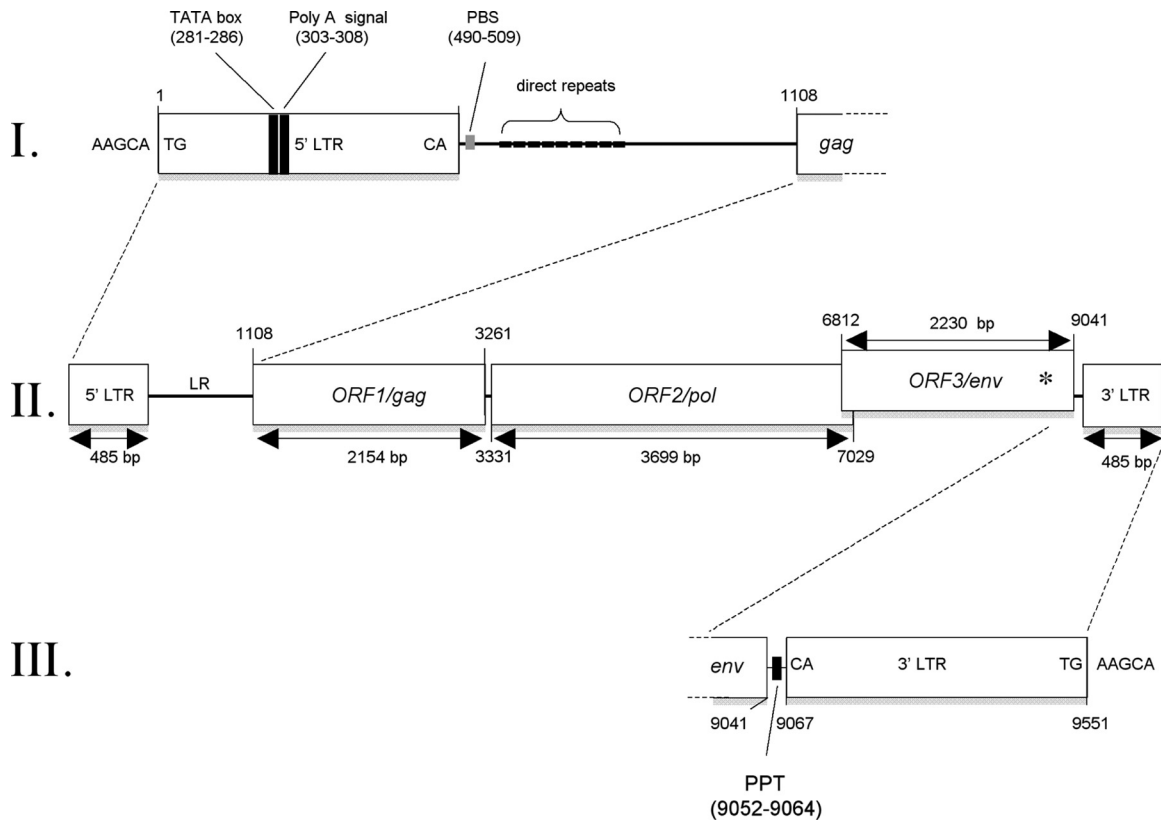


FIG. 2. Proviral organization of XTERV1. With 9,551 bp in length, the XTERV1 proviral genome is characterized by the classical genomic organization 5'-LTR-gag-pol-env-3'-LTR (II). The gag and pol genes are in the same frame and are predicted to be translated as a Gag-Pol polyprotein by suppression of a termination codon at the junction of both genes. env is in a different frame and overlaps the 3' end of pol. An asterisk symbolizes the frameshift introduced within the env sequence. The 5' LTR and the leader region are depicted in panel I with the positioning of the TATA box, the poly(A) signal, the PBS, and the direct repeats. The 3' LTR is located upstream of the polypurine tract (PPT) (III). XTERV1 is flanked by the 5-bp target site duplication 5'-AAGCA-3'.

occurring at position 54, suggesting a recent integration. We noticed that this genomic sequence was incomplete, due to the presence of a sequencing gap. Using *in silico* data and PCR amplifications, we reconstructed the whole genomic sequence of this retroviral genome, delineated by the LTR sequences (see Materials and Methods). The genomic assembly at this locus was further validated by PCR using a panel of 20 primer pairs and by sequencing (primers are provided in Data Set S1 in the supplemental material; GenBank accession no. HM765512, see Data Set S3 in the supplemental material).

The inspection of the coding potential of this DNA sequence indicated the presence of three long ORFs, including FR47/env (nucleotide [nt] positions 6817 to 9046 in the sequence in Data Set S3 in the supplemental material). The two other ORFs,

referred to as ORF1 and ORF2, showed significant similarities to and/or conserved functional motifs of the Gag and Pol proteins of known ERVs. Moreover, the classical architecture 5'-LTR-gag-pol-env-3'-LTR was fully supported. Therefore, we named this newly identified provirus XTERV1 (*Xenopus tropicalis* endogenous retrovirus 1). The architecture of the proviral sequence of XTERV1 is depicted in Fig. 2, and the complete annotated sequence is given in Data Set S3 in the supplemental material.

To question our model for whether XTERV1 is endogenous, i.e., encoded by the nuclear genome and inherited, we investigated the conservation of its genetic environment (insertion site) during transmission from the germ line to the offspring. To address this question, we performed PCR ampli-

FIG. 1. Characterization of *X. tropicalis* FR47-like protein as an envelope-like protein. (A) Sequence alignment of *R. sylvatica* FR47 ORF protein and *X. tropicalis* FR47-like protein. To maintain a proper alignment with the *R. sylvatica* FR47 ORF, a frameshift (shown as an X over a black background at position 669) was introduced within the *X. tropicalis* FR47-like protein, which is consequently annotated as a 743-aa-long protein. The first methionine defined elsewhere (43) at position 260 is boxed in black. SU and TM are delineated with the putative cleavage site (R, Y/H, K, R), shown in white letters over a blue background. The N-terminal hydrophobic signal peptide and the C-terminal transmembrane domain are shaded in blue and red, respectively. Putative glycosylation sites are boxed. Conservation of residues is according to the BLOSUM 62-aa matrix. (B) Multiple-sequence alignment of envelope proteins homologous to *R. sylvatica* FR47 and *X. tropicalis* FR47/Env proteins. The coloring scheme for the residues is according to the BLOSUM 62-aa matrix. The putative canonical cleavage sites (consensus R/K-X-R/K-R) are shaded in red. A consensus sequence is displayed below the alignment. Sequence information is available in Data Set S2 in the supplemental material.

fications on gDNA extracted from *X. tropicalis* sperm nuclei. A primer set was designed to amplify a 1,614-bp fragment spanning the end of *FR47/env* gene, the 3' LTR, and 54 bp of 3' flanking gDNA. We obtained a single amplicon, and its sequence was identical to the XTERV1 reference genomic sequence (data not shown). We conclude that the 9,551-bp-long XTERV1 proviral genome is a bona fide endogenous retroviral component of the *X. tropicalis* nuclear genome.

LTRs and leader region. Both LTR sequences display recognizable features of all ERVs, including the presence of the terminal sequences 5'-TG and 3'-CA flanking each LTR, as well as a 5-bp target site duplication (TSD), 5'-AAGCA-3', flanking the XTERV1 proviral sequence. Thus, the integration of XTERV1 followed the common rules of recombination used by retroviruses.

The XTERV1 5' LTR contains a presumptive TATA box signal (TATAAA; positions 281 to 286 in Fig. 2) and a polyadenylation signal at position 303 (Fig. 2). The 3' LTR follows a polypurine tract (GACTAAAAGGGGAT; positions 9052 to 9064 in Fig. 2). The leader region (LR) or 5' untranslated terminal region (5' UTR) of XTERV1, positioned before the *gag* ORF, is 622 bp long and contains nine short direct repeat sequences of 7 bp each (Fig. 2; see Data Set S3 in the supplemental material). The presence of several repeats is a common feature of most retroviruses identified in the genomes of vertebrates, such as zebra fish and *Xenopus* (32). The role(s) of these sequences remains largely unknown. The LR was also predicted to contain a potential primer-binding site (PBS) sequence adjacent to the 5' LTR (positions 490 to 509 in Fig. 2). Although the sequences of tRNAs from the *X. tropicalis* genome are not yet catalogued, the putative 20-bp XTERV1 PBS is closely related to the complementary sequence of the 3' end of the mouse leucine tRNA (CAG anticodon; 16/20 matches). This PBS sequence was also similar to that described in elements of human endogenous retrovirus with leucine tRNA primer (HERV-L) and its murine homolog, MuERV-L (4, 16). In conclusion, LTRs and LRs of XTERV1 contain the known functional elements required for retrovirus replication.

ORF1/Gag. The predicted ORF1 start codon is located at nucleotide position 1108 on the proviral genome (Fig. 2). ORF1 has a predicted size of 2,154 bp and encodes a putative 717-aa protein. BLAST searches using the ORF1 translated product did not reveal overt similarities with other Gag proteins. However, conserved motifs of the characteristic matrix (MA), capsid (CA), and nucleocapsid (NC) functional domains (see Data Set S3 in the supplemental material) were identified, including the myristylation sequence MGNKTS, a single NC motif known as Cys-His box (aa 648 to 661), and a glycine-arginine (GR)-rich region (aa 622 to 627). Moreover, the ORF1 sequence exhibits two of the canonical late domains, namely, LXPX_nL and PT/SAP, located at positions aa 155 (LYPNL motif) and aa 179 (PTAP motif), respectively. However, the L-domain PPPY is absent and may have been replaced by the functionally analogous proline-rich region PP-PPP at position aa 309. To conclude, XTERV1 ORF1 can be classified as a Gag protein due to the presence of typical Gag conserved motifs and its relative position following the 5' LTR and preceding the *pol* sequence.

ORF2/Pol. The second long ORF spans 3,699 bp and extends from nt 3331 to 7029 (Fig. 2). *gag* and *pol* are contiguous

in this region and are separated only by a termination codon. The *FR47/env* ORF lies in a different reading frame, and its 5' end overlaps the 3' end of the *pol* gene (spanning 218 bp of *pol* ORF). It is likely that the Gag and Pol proteins are synthesized as a large single polypeptide precursor via termination suppression, as described for the murine leukemia virus (62). The deduced Pol protein sequence resembles that of several retrovirus polymerases and contains structural domains arranged in a typical order: protease (PR), reverse transcriptase (RT), RNase H, and integrase (INT). Using a blastp search, the highest sequence identities recorded were of 34% and 35% for polyproteins in zebra finch (*Taeniopygia guttata*, GenBank accession no. XP_002195379) and opossum (*Monodelphis domestica*, GenBank accession no. XP_001377846) genomes, respectively. Therefore, XTERV1 ORF2 encodes a typical retroviral polyprotein with apparently intact coding capacity.

Genomic environment of XTERV1. We studied the genomic environment of the XTERV1 locus on scaffold_387. XTERV1 flanking sequences appeared to be devoid of exons. The two closest transcription units encode a piggyBac transposable element derived 4 (PGBD4) located approximately 100 kbp downstream of XTERV1 (in the opposite orientation) and the oligodendrocyte transcription factor 3 (*olig3*) located 270 kbp downstream of XTERV1 (scaffold_387, positions 854246 to 856071). Furthermore, the XTERV1 locus was found to be surrounded by different types of repetitive elements, such as an XLGST3 repeat and a fragment of TE_ORF_108 sequences.

Characterization of XTERV1-related sequences within *Xenopus* genomes. Since ERVs are retrotransposable elements, we first looked for XTERV1-related sequences in the *X. tropicalis* genome database. Using the complete sequence of XTERV1 as a query in blastn searches, we did not find any genomic sequence with an overall identity above 95% elsewhere in the genome, suggesting that full-length XTERV1 provirus is present at one copy per haploid genome. *X. tropicalis* genome database searches using the FR47/Env protein sequence as a query led us to the identification of 59 loci containing FR47/Env paralogs (see Data Set S4 in the supplemental material). For 11 loci, FR47/Env-like sequences were surrounded by LTRs, whereas 12 loci were found to contain solo LTR elements. Finally, 36 loci contained FR47/Env-related sequences but without any detectable LTR in the neighboring region (<20 kbp). Sequence analyses of all of these loci revealed significant differences with XTERV1, as each one carries multiple in-frame stop codons, frameshift mutations, and limited nucleotide sequence similarity. Thus, the XTERV1 scaffold_387 locus represents a particularly well-conserved element of an ERV family that populates the *X. tropicalis* genome.

To further characterize the XTERV1 family, we estimated the age of the integration events using LTR sequence divergence. If mutations occur independently in the two LTRs, the age of integration can be estimated from LTR nucleotide divergence. However, the precision of this estimate will depend on the proportion of interlocus gene conversion between different proviruses (29).

We compiled a set of LTR sequences and estimated the age of the oldest insertion to be 41 million years ago (mya; standard error, 8 mya), using a rate of nucleotide substitution (*r*) of 0.00103, as previously calculated for frog nuclear genes (17).

We documented further the evolutionary history of XTERV1 element insertion by inferring a phylogeny from 22 duplicated LTRs from 11 elements (see Data Set S5 in the supplemental material). LTRs from four elements were found to be unpaired, probably as a result of gene conversion or recombination with solo LTRs. The LTR tree highlights different subfamilies, suggesting several waves of integration events either by retrotransposition or by reinfection.

Next, we searched for XTERV1-related elements within the genome of the sibling species *X. laevis*. All PCR assays described above failed to amplify any fragment on *X. laevis* gDNA (data not shown). We identified a single sequence homologous to the XTERV1 *pol* gene, contained within the insert of the bacterial artificial chromosome clone CH219-98J22 (GenBank accession no. AC236462). Sequence alignment showed that a portion of the XTERV1 *pol* nucleotide sequence is similar to this BAC sequence, including a portion of the LTR, with 76% identity (see Data Set S6 in the supplemental material). We named this element XLERV1 (*Xenopus laevis* endogenous retrovirus 1), and sequence analysis revealed that it did not contain sequences similar to *FR47/env* (see Data Set S6 in the supplemental material). We validated the existence of the XLERV1 *pol* sequence by PCR on *X. laevis* gDNA.

Together with our estimation of an insertion time compatible with integration in the *Xenopodinae* ancestor, these data allowed us to conclude that the XTERV1 and XLERV1 sequences are homologous.

Copy number and chromosomal localization of XTERV1-related elements within *X. tropicalis* genome. To explore the chromosomal localization of XTERV1, we performed fluorescence *in situ* hybridization on metaphasic chromosome preparations from *X. tropicalis*. Four probes covering a large portion of the XTERV1 sequence were generated (probes *env*, LTR, *pol*, and *gag* in Fig. 3A).

Using the *env* probe, a unique and strong hybridization signal was visualized in the centromeric area of chr. 5 (19/23 spreads; Fig. 3B). Since the *olig3* gene was found to be located 270 kbp downstream of XTERV1 on scaffold_387, we took advantage of its recent characterization as a chr. 5 centromeric marker to perform FISH with the corresponding probe (34). Indeed, the *olig3* probe hybridized to the same centromeric region on chr. 5 (Fig. 3C). Since the genetic distance of *olig3* to the centromere was shown to be less than 0.9 cM (34), the location of XTERV1 (scaffold_387) is pericentromeric.

Using the *pol* probe, two to nine hybridization signals on interphase nuclei were observed (average, 5.6 signals per cell, i.e., 2.8 per haploid genome; $n = 45$; Fig. 3D). This result was consistent with that obtained *in silico* using the *pol* probe sequence as a query in blastn searches. Indeed, four loci located on two scaffolds, scaffold_387 and scaffold_559, could be identified (90% identity and coverage of >100 bp). The detection on metaphasic chromosomes was not systematic; nevertheless, two loci could be identified with certainty ($n = 31$): the *p* arm of the longest acrocentric chromosome, chr. 3 ($n = 8$, 25.8%), and the centromeric region of chr. 5 ($n = 8$, 25.8%) (Fig. 3E).

Similar results were obtained using the *gag* and LTR probes ($n = 43$), with FISH signals being observed on the two previously identified chromosomes, namely, chr. 3 ($n = 15$, 34.9%) and chr. 5 ($n = 16$, 37.2%) (Fig. 3F and G). An additional locus containing XTERV1 elements was identified on chr. 1 ($n = 16$,

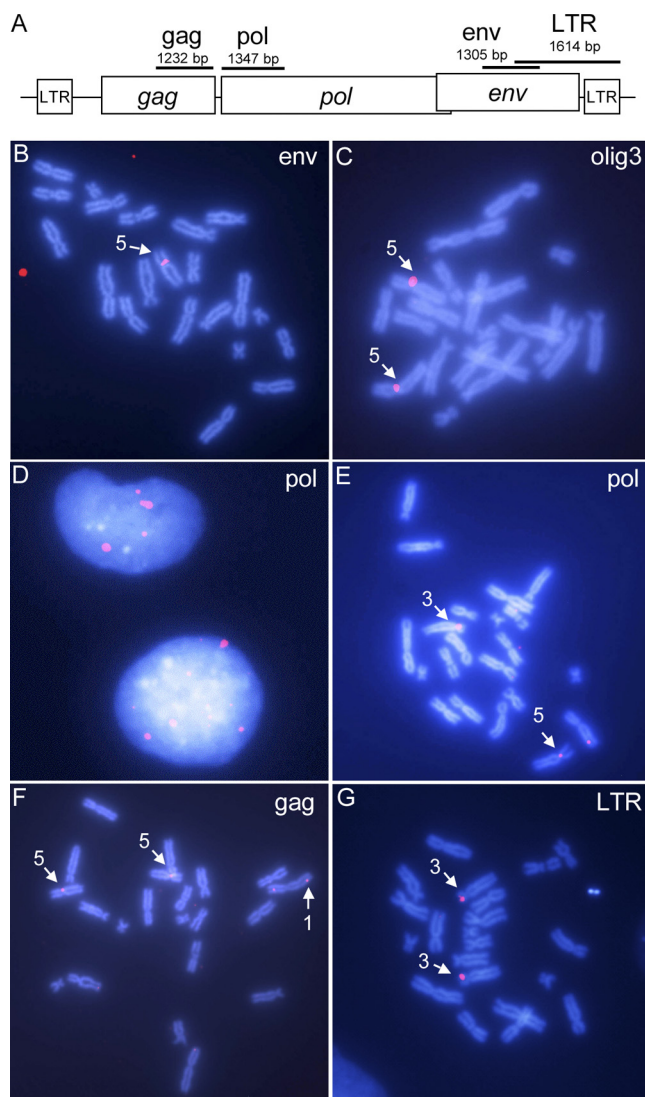


FIG. 3. Chromosomal localization of XTERV1 copies. (A) Schematic representation of XTERV1 genome showing the positions and the lengths of the probes used in this study. Representative images of metaphasic spreads after hybridization of the *env* probe (B), *olig3* probe (C), *pol* probe (E), *gag* probe (F), and LTR probe (G) in FISH experiments. The numbers next to the arrows are chromosome numbers. (D) Image of interphase nuclei after hybridization of the *pol* probe.

37.2%). In parallel, we observed that the hybridization signal obtained using the LTR probe on interphase nuclei was more abundant than that obtained using the *gag* and *pol* probes (data not shown). These results confirmed *in silico* data since 47 loci, distributed on 36 scaffolds, were identified in blastn searches using the LTR probe sequence as a query (90% identity, coverage of >100 bp). Thus, it is clear that solo LTR elements and LTRs are more highly represented than full-length elements within the *X. tropicalis* genome.

Taken together, both experimental and *in silico* data provide evidence that full-length XTERV1 elements exist at a low copy number (about 10 copies) within the *X. tropicalis* genome. These copies are interspersed, since three loci

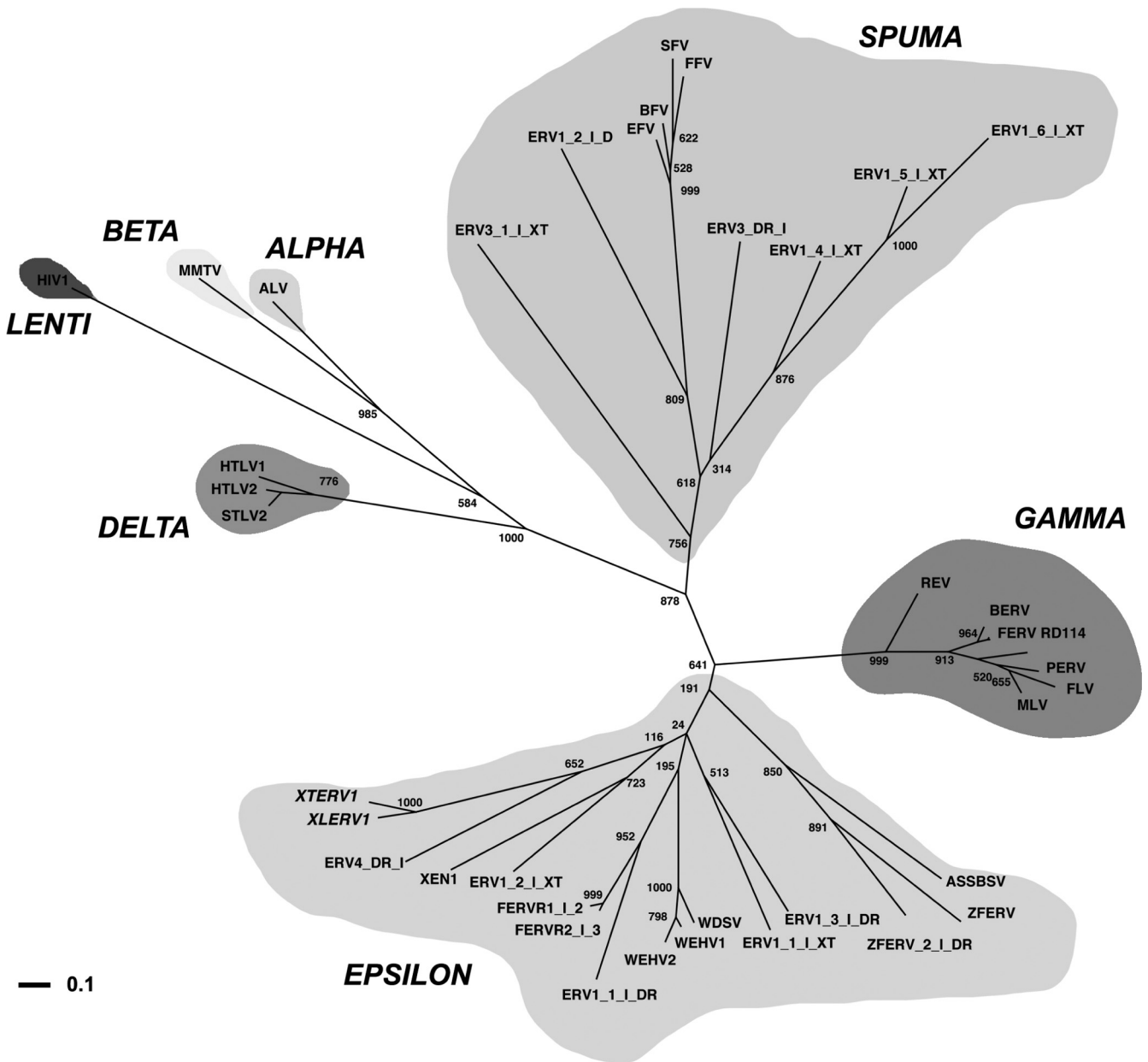


FIG. 4. Unrooted phylogenetic tree of representative retroviruses based on a MUSCLE multiple alignment of 40 amino acid sequences matching the Prosite reverse transcriptase domain (accession no. PS50878). This tree was calculated using a maximum-likelihood algorithm implemented in PhyML (25). Numbers above or below the branches indicate percent support for the nodes in distance bootstrap analysis (500 replicates). Horizontal branch lengths are proportional to the degree of amino acid substitutions per site. The log likelihood of this tree is $-9,652.52$. The retrovirus genera are indicated. Retroviruses are as follows: XTERV1 (accession no. HM765512), XLERV1 (accession no. AC236462.1), XEN1 (*Xenopus laevis*; accession no. AJ506107.1), ZFERV (*Danio rerio*; accession no. AF503912.1), baboon endogenous virus strain M7 (BERV; accession no. BAA89659.1), feline endogenous retrovirus (FERV; accession no. P31792.1), feline foamy virus (FFV; accession no. NP_056914.1), avian leukosis virus (ALV; accession no. NP_040550.1), equine foamy virus (EFV; accession no. NP_054716.1), feline leukemia virus (FLV; accession no. NP_047255.1), Gibbon ape leukemia virus (GALV; accession no. AAC80264.1), human immunodeficiency virus type 1 (HIV1; accession no. NP_057849.4), human T-lymphotropic virus type 1 (HTLV1; accession no. NP_057860.1), human T-lymphotropic virus type 2 (HTLV2; accession no. NP_041003.1), simian foamy virus (SFV; accession no. NP_044280.1), Moloney murine leukemia virus (MLV; accession no. P03355.4), mouse mammary tumor virus (MMTV; accession no. NP_955564.1), porcine endogenous retrovirus (PERV; accession no. CAC82505.2), simian T-lymphotropic virus type 2 (STLV2; accession no. NP_056907.1), WDSV (accession no. NP_045937.1), WEHV1 (accession no. AF133051_3), WEHV2 (accession no. AAC59311.1), reticuloendotheliosis virus (REV; accession no. AAZ57418.1), bovine foamy virus (BFV; accession no. NP_044929.1), ASSBSV (accession no. YP_443922.1), and RD114 retrovirus (RD114; accession no. YP_001497148.1). All sequences prefixed with ERV or FERV are from Replibase (31), including ZFERV_2_I_DR, with the organisms of origin given the suffixes *Takifugu rubripes* (FERV), *Danio rerio* (DR), or *Xenopus tropicalis* (XT).

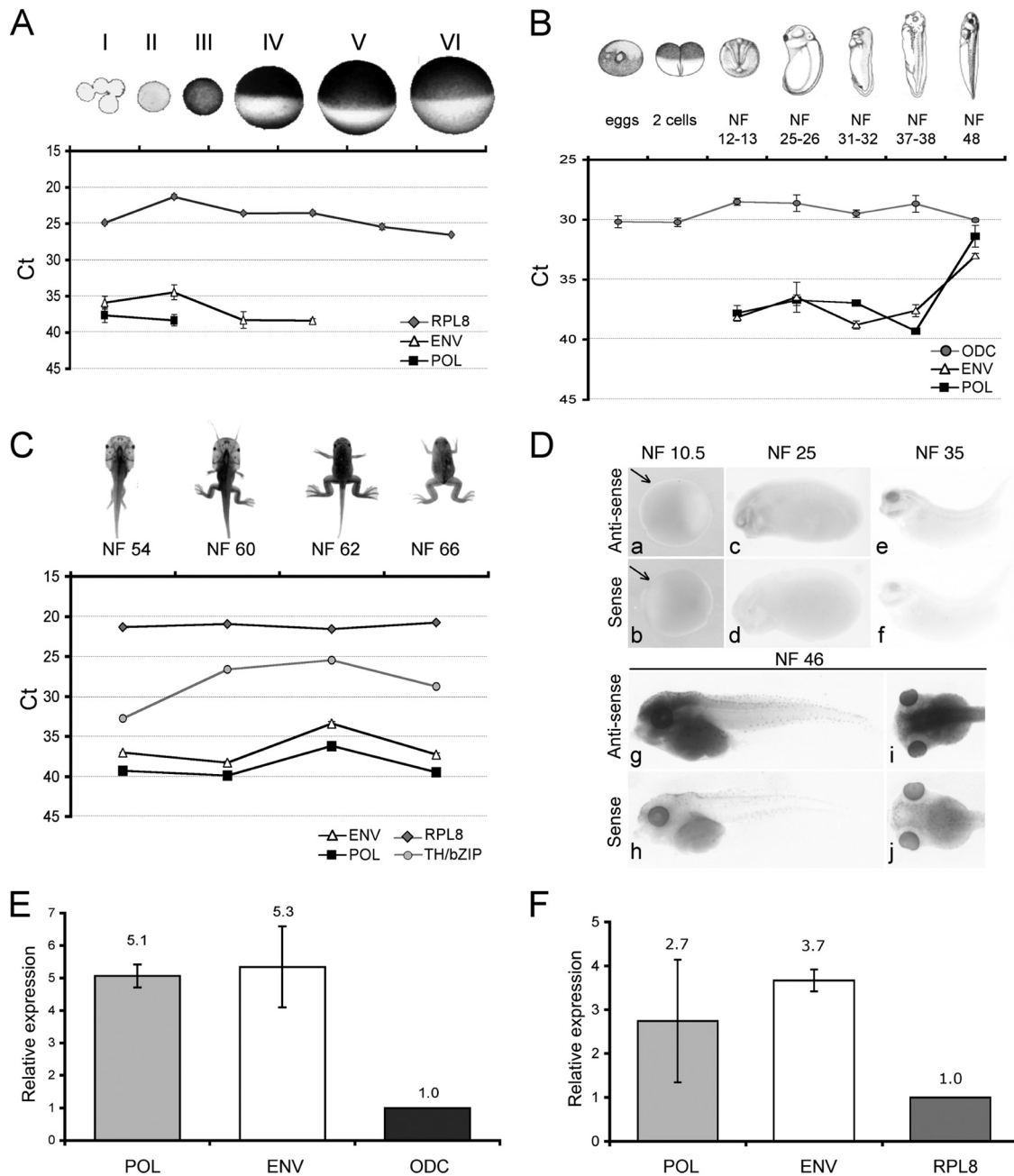


FIG. 5. Expression of XTERV1 genes. Temporal expression profiles of XTERV1 *pol* and *FR47/env* genes during oogenesis (A), early development (B), and premetamorphic and metamorphic stages (C). Amounts of cDNA were used to monitor the levels of transcripts during ontogenesis. The results are expressed as the mean C_T values (in reverse on the left axis to represent mRNA abundance) as a function of the developmental stage (x axis). Ornithine decarboxylase (ODC) and ribosomal protein L8 (RPL8) transcripts are typically used to normalize RT-qPCR results of developmental time series in *Xenopus* (53). Error bars represent standard deviations among biological replicates. *TH/bZip* transcript levels were measured since its levels are under TH regulation during metamorphosis. (D) Spatial expression profile of *FR47/env* transcripts using whole-mount in situ hybridization experiments at the indicated stages. WISH signals were not observed when the control sense probe was used. (a and b) Lateral view of stage NF10.5; the arrows indicate the blastopore lip; (c and d) lateral view of stage NF25; (e and f) lateral view of stage NF35; (g and h) lateral view of stage NF46; (i and j) dorsal view of stage 46. XTERV1 stress response. (E and F) Expression analysis of *pol* and *FR47/env* genes in cells submitted to serum starvation (E) and UV exposure (F). The relative expression levels of the XTERV genes were normalized to the expression levels of ornithine decarboxylase transcripts and ribosomal protein L8. The error bars indicate the standard deviations between biological replicates.

could be assigned to chr. 1, 3, and 5. A single complete full-length XTERV1 element with classical genomic organization and apparently intact ORFs for the *gag* and *pol* genes was identified on scaffold_387 and mapped to the centromeric region of chr. 5.

Phylogenetic analyses. To investigate the phylogenetic relationships of XTERV1 and XLERV1 with other known retroviruses, we constructed a multiple alignment of retroviral reverse transcriptase sequences from representative members of all seven retroviral genera. As shown in Fig. 4, the resulting

phylogenetic tree grouped XTERV1 and XLERV1 together with a zebra fish (*Danio rerio*) ERV (Repbase ERV4-DR-I). This group is placed within the *Epsilonretrovirus* genus but separately from the other piscine ERVs, such as walleye-derived retroviruses, zebra fish endogenous retrovirus 1 (ZFERV1), and ZFERV2. We conclude that XTERV1 is related to the *Epsilonretrovirus* genus.

Expression of *pol* and *FR47/env* genes throughout *X. tropicalis* development. Database searches highlighted expressed sequence tags derived from various tissues and related to the *gag*, *pol*, and *FR47/env* XTERV1 genes. In *Xenopus*, previous analyses have shown that abundant levels of FR47 transcripts were present in embryos, in metamorphic limbs, as well as in some adult tissues (22). Moreover, the presence of a TATA box and a polyadenylation signal within the LTRs strongly suggests that XTERV1 can be transcribed. To gain insights into the expression of XTERV1 during *X. tropicalis* development, primers specific for the *pol* and *FR47/env* genes were designed to perform real-time PCR on reverse transcription products.

At the early stages of oogenesis, both the *pol* and *FR47/env* amplicons revealed low levels of XTERV1 transcripts. As oocyte maturation proceeded, XTERV1 transcripts were no longer detected (Fig. 5A). *pol* transcripts were detected in RNAs from stages I and II, whereas *FR47/env* transcripts were observed until stage IV. During early development, neither transcript was detected in unfertilized eggs and two-cell-stage embryos, in line with the previous results for oogenesis. XTERV1 transcripts were not maternally supplied and were detected at the early neurula stage (NF12-NF13), after the initiation of zygotic transcription that occurs at the midblastula transition in *Xenopus*. The levels of XTERV1 transcripts remained stable in embryos from the neurula stage to the late-tail-bud stage (NF37-NF38) and then increased significantly to reach nearly the same amounts as *odc* transcripts at tadpole stage 48. These results were confirmed by WISH experiments. To avoid confusion due to the presence of the multiple copies of XTERV1, we documented only the spatial expression of *FR47/env*-containing transcripts, since the *env* probe was found to specifically target the expected complete XTERV1 copy (scaffold_387) in FISH analysis (Fig. 3B). At the gastrulation stage, stage NF10.5, *FR47/env* transcripts were not detected, in agreement with RT-qPCR data (Fig. 5D, panel a). By stages NF25 and NF35, *FR47/env* expression was well-defined within the head region, including the developing eye (Fig. 5D, panels c and e). As development proceeds, its expression strongly increased (Fig. 5D, panels g and i) and the signal appeared prominent in the brain and eyes (stage NF46). However, there was no strong tissue or cell specificity of expression.

During metamorphosis, *FR47/env* and *pol* transcripts were found to be transiently upregulated between stages 60 and 62, with 30- and 13-fold differences, respectively (Fig. 5C). In amphibians, the metamorphic climax occurs at stage 62, when the TH concentration peaks and TH target genes, such as *TH/bzip*, are robustly induced (12). Indeed, *TH/bzip* expression increased 154-fold from the beginning (NF54) to the climax (NF62) of metamorphosis (Fig. 5C). After the climax of metamorphosis, *TH/bzip*, *FR47/env*, and *pol* expression levels decreased up to stage 66 (young juvenile) to reach the levels observed at premetamorphic stages. However, XTERV1 transcription was not in-

duced by the TH signaling pathway, since no significant change for *pol* and *FR47/env* transcripts was observed in RNA extracted from cells with or without an exogenous supply of physiological concentrations of TH (data not shown).

In conclusion, the *pol* and *FR47/env* genes harbor similar dynamic expression profiles, and XTERV1 gene expression is regulated during *X. tropicalis* development.

Regulation of XTERV1 genes under cellular stress conditions. Several environmental clues (UV exposure, chemicals, drugs, and stress) or endogenous clues (such as hormonal balance) have been shown or suggested to influence the expression of endogenous retroviruses (51, 53). Therefore, we quantified the expression levels of XTERV1 genes following various cellular stresses to characterize their regulation in comparison with the stress responses described for *R. sylvatica* FR47 (43).

We first investigated the effects of a cold exposure on XTERV1 gene expression, since *FR47* transcripts are upregulated after freezing of *R. sylvatica*. However, no significant alteration in the transcription profiles of XTERV1 *pol* and *FR47/env* transcripts between control and cold-shocked cells or in *X. tropicalis* tadpoles subjected to cold conditions was noticed (data not shown). Next, cells were subjected to a metabolic stress. For this purpose, we applied serum starvation and compared *pol* and *FR47/env* RNA expression profiles under these culture conditions. Under starvation conditions, a significant increase in *pol* and *FR47/env* mRNA levels was observed, with the changes being 5.07- and 5.34-fold, respectively (Fig. 5E). Thus, it appears that XTERV1 transcripts are upregulated in cells under serum deprivation.

Finally, we exposed Speedy cells to UV radiation to determine if a different type of stress would affect XTERV1 gene expression (Fig. 5F). We observed 2.74- and 3.67-fold increases in *pol* and *FR47/env* transcript levels upon UV irradiation.

Taken together, we found that XTERV1 expression was not affected by cold shock but was upregulated following UV irradiation (cytotoxic stress) and serum starvation (metabolic stress). These results support a fine regulation of XTERV1 transcription, which appears to respond specifically to certain types of stresses.

DISCUSSION

Genomic characteristics of XTERV1. A first view of the genome landscape of ERVs in an amphibian was obtained upon the recent completion of the *X. tropicalis* genome (26). ERVs belonging to classes I and III could be identified using the RepeatMasker program, and their abundances were estimated to be 0.02 and 0.10%, respectively. Therefore, retroviral amplification did not substantially impact the genome of *Xenopus*. In another study, Basta and colleagues mentioned the existence of such *X. tropicalis* ERVs in relation to teleost fish retroviruses but did not provide details (1).

XTERV1 elements are interspersed within the *X. tropicalis* genome but are not abundant. We detected a single complete full-length XTERV1 element (mapped at the centromeric region of chr. 5) accompanied by less than 10 full-length defective copies without any coding capacity (Fig. 3), two of them being localized to chr. 3 and chr. 1. We estimated the time of infection to be 41 mya, meaning that it likely happened in the *Xenopodinae* ancestor. This approximate age is coherent with

our finding of a related XLERV1 element in *Xenopus laevis*. We detected several waves of insertion, including very recent ones, since the scaffold_387 element is complete with 99.8% identical LTRs. This suggests the persistence up to the present day of a pool of an active element(s) or of recurrent infections.

Our study was first initiated to characterize transcripts highly expressed during metamorphosis in *X. tropicalis* (22). The conceptual translation of one candidate was found to be closely related to the freeze-response FR47 protein in the wood frog (*R. sylvatica*) (43) but was also similar to the Env proteins of several mammalian ERVs (Fig. 1). Further investigations of the *X. tropicalis* genomic locus encoding this transcript allowed us to report the first complete full-length sequence of a simple endogenous retrovirus within the *X. tropicalis* genome, named XTERV1 (Fig. 2). XTERV1 harbors the universal ERV structure with three long ORFs encoding *gag*, *pol*, and *env* flanked by LTRs with the typical genetic order 5'-LTR-*gag-pol-env*-3'-LTR. The LTRs contained expected and intact regulatory sequences, such as a TATA box and a polyadenylation signal. Furthermore, the almost perfect identity (only one substitution) between the two LTRs suggests that at least one XTERV1 element was inserted quite recently. The conceptual translation of the *gag* and *pol* region suggested the possible expression of a Gag-Pol polyprotein precursor through suppression of an amber termination codon, similar to what is observed in walleye-derived retroviruses. However, the *env* gene required an arbitrary frameshift just upstream the transmembrane region to maintain proper alignment to other known Env proteins, suggesting that XTERV1 either is unable to encode an Env protein, produces a secreted protein, or uses an alternative yet unidentified mechanism to produce a membrane-anchored Env. Similar examples have been found in the case of the EnvR from *env3* and *EnvF(c)2*, and protein products could be obtained using *in vitro* transcription-translation assays (15, 18). We made an exhaustive survey of the *X. tropicalis* genome to identify XTERV1 *env* paralogs, and we did not detect a single one without mutations.

Phylogenetic analysis based on Pol protein alignments (Fig. 4) revealed that XTERV1 is related to the group of *Epsilonretrovirus*-related ERVs from piscine hosts, including ZFERV1 and walleye-derived retroviruses (1, 28, 40, 41). XTERV1 shares several features common to most piscine retroviruses. It exhibits a long leader region of 622 bp with direct repeats, as seen for the Xen-1, ZFERV1, and the exogenous Atlantic salmon swim bladder sarcoma virus (ASSBSV) (32, 47, 52). Another important feature common to piscine retroviruses, including walleye epidermal hyperplasia virus type 1 (WEHV1), WHEV2, and members of the gammaretrovirus group, is that XTERV1 contains a single Cys-His box in the NC protein, while retroviruses of other genera have two such motifs (40).

XTERV1 exhibits singular features in comparison to other known epsilonretroviruses since it uses a tRNA^{Leu} binding site for replication. Indeed, the tRNA is specific to the virus genus and other members of *Epsilonretrovirus* group such as Xen-1 use a tRNA^{Lys2} (32). The XTERV1 PBS sequence is similar to that of HERV-L and MuERV-L (4, 16). Another interesting functional difference with piscine ERV resides in the fact that XTERV1 is a simple ERV and does not contain any accessory genes, such as a homologue of the cellular 2',3'-cyclic nucleo-

tide 3'-phosphodiesterase gene (CNPase) found in three different positions in various fish ERVs genomes such as walleye dermal sarcoma virus (WDSV), WEHV1, WEHV2, ZFERV, and snakehead retrovirus (1). All the piscine ERVs are also characterized by additional short ORFs of about 100 aa located within the LR and positioned before the *gag* sequence (32, 52). The XTERV1 LR does not exhibit such putative ORFs.

Regulation of XTERV1 expression. We report that XTERV1 genes are transcribed in a regulated manner in various tissues and organs during *X. tropicalis* development (Fig. 5). The most remarkable feature of XTERV1 gene expression remains the upregulation observed at the metamorphic climax, which was consistent with *in silico* data (22). Such an expression profile is reminiscent of the expression profiles of other genes induced during metamorphosis, a process known to be under endocrine control by TH in amphibians. However, we did not observe a correlation between TH levels and XTERV1 mRNA levels. Nevertheless *pol* and *FR47/env* might be upregulated during metamorphosis either as a consequence of the morphological changes and associated cellular differentiation events observed at this stage or as a response to other endocrine signals such as corticosteroids (stress hormones) that are known to modulate the actions of TH at the time of metamorphosis (35).

In *Xenopus*, other TEs are expressed during development and exhibit restricted expression profiles (24, 33). However, the expression profiles of TEs during development are different from each other. For example, *Xretpos* retrotransposon expression is zygotically activated, restricted to ventroposterior-specific regions, and induced by ventralizing manipulation, such as UV irradiation (53). Thus, the expression profile of XTERV1 differs from what has been reported to date concerning other TEs in *Xenopus*, and therefore, we can postulate that specific regulation cues might drive XTERV1 transcription during *Xenopus* development.

Syncytin proteins constitute a strong and well-documented case of retrovirus-derived proteins with crucial biological functions in mammals (7, 10, 44). Similarly, the envelope protein of the endogenous Jaagsiekte sheep retroviruses (enJSRVs) plays fundamental roles in placental morphogenesis and mammalian reproduction (20). Analogous cases could exist in nonmammalian vertebrates. Previous reports have shown that ERVs can display variable and specific spatiotemporal expression patterns. In larval and adult zebra fish, ZFERV is predominantly expressed in the thymus (52). In chickens, OVEX1 is transcribed in gonads with a sex-dependent and left-right asymmetrical pattern (13). OVEX1 retroviral proteins could be actors or witnesses in the processes of ovarian development and therefore could fulfill a biological role(s) (13).

McNally and colleagues described the freeze-inducible gene *FR47*, which encodes a protein expressed in the livers of freeze-tolerant anurans and which is upregulated during freezing and thawing (43). Moreover, *FR47* transcript levels were increased following freezing, anoxia, and dehydration stresses. The authors suggest that *FR47* function is important for freeze survival. In this study, we brought evidence that *FR47* mRNA encodes an envelope protein, derived from a retroviral element within the *R. sylvatica* genome. In light of these findings on *FR47* in *R. sylvatica* and on XTERV1 expression in *X. tropi-*

calis, further studies are required to investigate the potential role(s) of these particular retroviral elements in the regulation of cellular and developmental processes in these vertebrate models.

ACKNOWLEDGMENTS

We are grateful to Yves Bigot and to Andrew Tindall for critical reading of the manuscript.

This work was supported by Genopole, CNRS, the University of Evry, and the Conseil General de l'Essonne (ASTRE T-REX). We acknowledge the support of the CERFAP facility from Genopole.

REFERENCES

- Basta, H. A., S. B. Cleveland, R. A. Clinton, A. G. Dimitrov, and M. A. McClure. 2009. Evolution of teleost fish retroviruses: characterization of new retroviruses with cellular genes. *J. Virol.* **83**:10152–10162.
- Baust, C., W. Seifarth, U. Schon, R. Hehlmann, and C. Leib-Mosch. 2001. Functional activity of HERV-K-T47D-related long terminal repeats. *Virology* **283**:262–272.
- Belshaw, R., and A. Katzourakis. 2005. BlastAlign: a program that uses blast to align problematic nucleotide sequences. *Bioinformatics* **21**:122–123.
- Benit, L., et al. 1997. Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. *J. Virol.* **71**:5652–5657.
- Benit, L., P. Dessen, and T. Heidmann. 2001. Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J. Virol.* **75**:11709–11719.
- Best, S., P. Le Tissier, G. Towers, and J. P. Stoye. 1996. Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* **382**:826–829.
- Blaise, S., N. de Parseval, L. Benit, and T. Heidmann. 2003. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U. S. A.* **100**:13013–13018.
- Blikstad, V., F. Benachenhou, G. O. Sperber, and J. Blomberg. 2008. Evolution of human endogenous retroviral sequences: a conceptual account. *Cell. Mol. Life Sci.* **65**:3348–3365.
- Blomberg, J., F. Benachenhou, V. Blikstad, G. Sperber, and J. Mayer. 2009. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* **448**:115–123.
- Blond, J. L., et al. 2000. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J. Virol.* **74**:3321–3329.
- Bronchain, O. J., et al. 2007. The olig family: phylogenetic analysis and early gene expression in *Xenopus tropicalis*. *Dev. Genes Evol.* **217**:485–497.
- Brown, D. D., et al. 1996. The thyroid hormone-induced tail resorption program during *Xenopus laevis* metamorphosis. *Proc. Natl. Acad. Sci. U. S. A.* **93**:1924–1929.
- Carre-Eusebe, D., N. Coudouel, and S. Magre. 2009. OVEX1, a novel chicken endogenous retrovirus with sex-specific and left-right asymmetrical expression in gonads. *Retrovirology* **6**:59.
- Cohen, C. J., W. M. Lock, and D. L. Mager. 2009. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. *Gene* **448**:105–114.
- Cohen, M., M. Powers, C. O'Connell, and N. Kato. 1985. The nucleotide sequence of the env gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. *Virology* **147**:449–458.
- Cordonnier, A., J. F. Casella, and T. Heidmann. 1995. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. *J. Virol.* **69**:5890–5897.
- Crawford, A. J. 2003. Relative rates of nucleotide substitution in frogs. *J. Mol. Evol.* **57**:636–641.
- de Parseval, N., V. Lazar, J. F. Casella, L. Benit, and T. Heidmann. 2003. Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins. *J. Virol.* **77**:10414–10422.
- Dimcheff, D. E., M. Krishnan, and D. P. Mindell. 2001. Evolution and characterization of tetraonine endogenous retrovirus: a new virus related to avian sarcoma and leukosis viruses. *J. Virol.* **75**:2002–2009.
- Dunlap, K. A., et al. 2006. Endogenous retroviruses regulate periimplantation placental growth and differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **103**:14390–14395.
- Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**:1792–1797.
- Fierro, A. C., et al. 2007. Exploring nervous system transcriptomes during embryogenesis and metamorphosis in *Xenopus tropicalis* using EST analysis. *BMC Genomics* **8**:118.
- Gifford, R., and M. Tristem. 2003. The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* **26**:291–315.
- Greene, J. M., H. Otani, P. J. Good, and I. B. Dawid. 1993. A novel family of retrotransposon-like elements in *Xenopus laevis* with a transcript inducible by two growth factors. *Nucleic Acids Res.* **21**:2375–2381.
- Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**:696–704.
- Hellsten, U., et al. 2010. The genome of the western clawed frog *Xenopus tropicalis*. *Science* **328**:633–636.
- Herniou, E., et al. 1998. Retroviral diversity and distribution in vertebrates. *J. Virol.* **72**:5955–5966.
- Holzschu, D. L., et al. 1995. Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. *J. Virol.* **69**:5320–5331.
- Hughes, J. F., and J. M. Coffin. 2001. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. *Nat. Genet.* **29**:487–489.
- Jern, P., and J. M. Coffin. 2008. Host-retrovirus arms race: trimming the budget. *Cell Host Microbe* **4**:196–197.
- Jurka, J., et al. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**:462–467.
- Kambol, R., P. Kabat, and M. Tristem. 2003. Complete nucleotide sequence of an endogenous retrovirus from the amphibian, *Xenopus laevis*. *Virology* **311**:1–6.
- Kay, B. K., M. Jamrich, and I. B. Dawid. 1984. Transcription of a long, interspersed, highly repeated DNA element in *Xenopus laevis*. *Dev. Biol.* **105**:518–525.
- Khokha, M. K., et al. 2009. Rapid gynogenetic mapping of *Xenopus tropicalis* mutations to chromosomes. *Dev. Dyn.* **238**:1398–1446.
- Krain, L. P., and R. J. Denver. 2004. Developmental expression and hormonal regulation of glucocorticoid and thyroid hormone receptors during metamorphosis in *Xenopus laevis*. *J. Endocrinol.* **181**:91–104.
- Kroll, K. L., and E. Amaya. 1996. Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development* **122**:3173–3183.
- Krylov, V., T. Tlapakova, and J. Macha. 2007. Localization of the single copy gene *Mdh2* on *Xenopus tropicalis* chromosomes by FISH-TSA. *Cytogenet. Genome Res.* **116**:110–112.
- Lander, E. S., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.
- Landry, J. R., A. Rouhi, P. Medstrand, and D. L. Mager. 2002. The *Opitz syndrome* gene *Mid1* is transcribed from a human endogenous retroviral promoter. *Mol. Biol. Evol.* **19**:1934–1942.
- LaPierre, L. A., D. L. Holzschu, P. R. Bowser, and J. W. Casey. 1999. Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: evidence for a gene duplication. *J. Virol.* **73**:9393–9403.
- LaPierre, L. A., D. L. Holzschu, G. A. Wooster, P. R. Bowser, and J. W. Casey. 1998. Two closely related but distinct retroviruses are associated with walleye discrete epidermal hyperplasia. *J. Virol.* **72**:3484–3490.
- Livak, K. J., and T. D. Schmittgen. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25**:402–408.
- McNally, J. D., C. M. Sturgeon, and K. B. Storey. 2003. Freeze-induced expression of a novel gene, *fr47*, in the liver of the freeze-tolerant wood frog, *Rana sylvatica*. *Biochim. Biophys. Acta* **1625**:183–191.
- Mi, S., et al. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**:785–789.
- Nieuwkoop, P. D., and F. Faber. 1994. Normal table of *Xenopus laevis* (Daudin). A systematic and chronological survey of the development from fertilized egg till the end of metamorphosis. Garland, New York, NY.
- Nolan, T., R. E. Hands, and S. A. Bustin. 2006. Quantification of mRNA using real-time RT-PCR. *Nat. Protoc.* **1**:1559–1582.
- Paul, T. A., et al. 2006. Identification and characterization of an exogenous retrovirus from Atlantic salmon swim bladder sarcomas. *J. Virol.* **80**:2941–2948.
- Pollet, N., et al. 2005. An atlas of differential gene expression during early *Xenopus* embryogenesis. *Mech. Dev.* **122**:365–439.
- Rasar, M. A., and S. R. Hammes. 2006. The physiology of the *Xenopus laevis* ovary. *Methods Mol. Biol.* **322**:17–30.
- Samuelson, L. C., K. Wiebauer, C. M. Snow, and M. H. Meisler. 1990. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol. Cell. Biol.* **10**:2513–2520.
- Serafino, A., et al. 2009. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Exp. Cell Res.* **315**:849–862.
- Shen, C. H., and L. A. Steiner. 2004. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. *J. Virol.* **78**:899–911.
- Shim, S., S. K. Lee, and J. K. Han. 2000. A novel family of retrotransposons in *Xenopus* with a developmentally regulated expression. *Genesis* **26**:198–207.
- Sindelka, R., Z. Ferjentsik, and J. Jonak. 2006. Developmental expression profiles of *Xenopus laevis* reference genes. *Dev. Dyn.* **235**:754–758.

55. **Sinzelle, L., et al.** 2006. Generation of transgenic *Xenopus laevis* using the Sleeping Beauty transposon system. *Transgenic Res.* **15**:751–760.
56. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**:1596–1599.
57. **Ting, C. N., M. P. Rosenberg, C. M. Snow, L. C. Samuelson, and M. H. Meisler.** 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* **6**:1457–1465.
58. **Tristem, M., E. Herniou, K. Summers, and J. Cook.** 1996. Three retroviral sequences in amphibians are distinct from those in mammals and birds. *J. Virol.* **70**:4864–4870.
59. **Wang-Johanning, F., et al.** 2007. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *Int. J. Cancer* **120**: 81–90.
60. **Waterston, R. H., et al.** 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**:520–562.
61. **Xu, Z., and H. Wang.** 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**:W265–W268.
62. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. U. S. A.* **82**:1618–1622.