JOURNAL OF VIROLOGY, Mar. 2006, p. 2941–2948 Vol. 80, No. 6
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.6.2941–2948.2006

Identification and Characterization of an Exogenous Retrovirus from Atlantic Salmon Swim Bladder Sarcomas

Thomas A. Paul,[1] Sandra L. Quackenbush,[2] Claudia Sutton,[1] Rufina N. Casey,[1] Paul R. Bowser,[1] and James W. Casey[1]*

Department of Microbiology and Immunology, Cornell University, Ithaca, New York 14853,[1] and Department of Microbiology, Immunology, and Pathology, Colorado State University, Fort Collins, Colorado 80523[2]

Received 24 August 2005/Accepted 15 December 2005

A novel piscine retrovirus has been identified in association with an outbreak of leiomyosarcoma in the swim bladders of Atlantic salmon. The complete nucleotide sequence of the Atlantic salmon swim bladder sarcoma virus (SSSV) provirus is 10.9 kb in length and shares a structure and transcriptional profile similar to those of murine leukemia virus-like simple retroviruses. SSSV appears unique to simple retroviruses by not harboring sequences in the Atlantic salmon genome. Additionally, SSSV differs from other retroviruses in potentially utilizing a methionine tRNA primer binding site. SSSV-associated tumors contain high proviral copy numbers (greater than 30 per cell) and a polyclonal integration pattern. Phylogenetic analysis based on reverse transcriptase places SSSV with zebrafish endogenous retrovirus (ZFERV) between the *Gammaretrovirus* and *Epsilonretrovirus* genera. Large regions of continuous homology between SSSV and ZFERV Gag, Pol, and Env suggest that these viruses represent a new group of related piscine retroviruses.

The etiological relationship between retroviruses and cancer is well established in mammalian systems. Retroviruses alter cell proliferation through the capture and deregulated expression of cellular proto-oncogenes, integration near or within the coding or regulatory regions of proto-oncogenes, or production of oncogenic viral gene products (10). In lower vertebrates, the role of retroviruses in tumor diseases has been established primarily through electron microscopic observations and detection of reverse transcriptase (RT) activity (12). The most thoroughly studied lower vertebrate retroviruses are the complex piscine epsilonretroviruses: walleye dermal sarcoma virus (WDSV) and walleye epidermal hyperplasia virus type 1 (WEHV-1) and type 2 (WEHV-2). Molecular characterization of these viruses from their associated skin proliferative diseases has provided insights into their mechanisms of pathogenesis. In particular, all three viruses share a complex structure and express distant cyclin homologs presumably involved in tumorigenesis (18, 19, 25, 34). Interestingly, WDSV- and WEHV-associated neoplasms have been documented to appear and regress seasonally, suggesting a complex regulation of tumor progression in these systems (11).

An outbreak of neoplastic disease of the swim bladders of Atlantic salmon (*Salmo salar*) with a suspected viral etiology was first reported at a Scottish commercial marine fish farm in 1975 (30). Affected salmon were in poor physical condition and displayed swollen abdomens that upon necropsy were found to contain multinodular masses of neoplastic cells occupying external and internal surfaces of the swim bladder. Severely affected fish had tumors running the entire length of the swim bladder that often protruded from the organ and occupied a large portion of the abdominal cavity. Histologically, tumors arose at the junction of the inner smooth muscle layer and the areolar tissue zone of the swim bladder and consisted of interlacing bundles of highly mitotic and well-differentiated spindle cells. Tumors were classified as leiomyosarcomas or fibrosarcomas (30). Additional examination of swim bladder sarcoma tissue by electron microscopy revealed the presence of numerous electron-dense, spheroidal bodies resembling retroviral particles (9).

A second outbreak of salmon swim bladder sarcoma was reported in juvenile salmon collected from the Pleasant River of Maine in 1996 and housed at the North Attleboro National Fish Hatchery (North Attleboro, MA) (4). Diseased fish exhibited multinodular masses on or replacing the swim bladder that histologically resembled previous reports of the disease (30). Affected fish were lethargic and displayed multifocal hemorrhages on the body and fins. Mortality among the affected lots of salmon peaked in late spring of 1998 at 35% of the population (4).

Based on the previous observation of retrovirus-like particles in swim bladder sarcoma tumors, degenerate PCR was utilized to amplify a tumor-associated exogenous retroviral sequence. By PCR, up to 60% of the affected salmon held at the North Attleboro National Fish Hatchery were infected with the retrovirus, whereas healthy fish had no detectable retroviral sequences (4). We report the cloning, complete nucleotide sequence, and transcriptional profile of Atlantic salmon swim bladder sarcoma virus (SSSV) from swim bladder sarcomas. Phylogenetic analysis of *pol* sequences suggests that SSSV is most closely related to a recently sequenced zebrafish endogenous retrovirus (ZFERV) (41) and that these viruses represent a new group of piscine retroviruses.

MATERIALS AND METHODS

Collection of tumor samples. Atlantic salmon used in this study were collected as juveniles from the Pleasant River of Maine between 1996 and 1997 and reared for brood stock at the North Attleboro National Fish Hatchery (North Attleboro, MA). Salmon held at North Attleboro exhibited clinical pathologies including

* Corresponding author. Mailing address: Department of Microbiology and Immunology, C4137 VMC, Cornell University, Ithaca, NY 14853. Phone: (607) 253-3412. Fax: (607) 253-3384. E-mail: jwc3@cornell.edu.

skin discoloration and hemorrhages on the fins and body that were first observed in 1997. In October 1997, tumors were first observed on swim bladders. Standard testing of affected salmon by the U.S. Fish and Wildlife Service Laboratories (Lamar, PA) for parasitic, bacterial, and viral pathogens was negative. In June 1998, frozen specimens of tumor tissue were shipped to Cornell University (Ithaca, NY) for evaluation of a potential retroviral etiology.

Identification of SSSV. Total RNA was purified from tumor and normal tissues (20 to 60 mg) using the RNAzol extraction procedure (Tel-test Laboratories, Friendswood, TX). Identification of retroviral *pol* sequences was accomplished by degenerate RT-PCR employing forward and reverse degenerate primers representing conserved amino acid motifs in polymerase VLPQG (5′-TAC CAGTGGAATGTTCTACCNCARGGN-3′) and YMDD (5′-ATCAGATCC TACTAACDRTCRTCCATRTA-3′), respectively. One microgram of total RNA was treated with RNase-free DNase I (Invitrogen, Carlsbad, CA) for 1 h at 37°C followed by DNase I inactivation with 2.5 mM EDTA for 10 min at 65°C. cDNA prepared from 1 μg of DNase I-treated RNA using the YMDD primer and Superscript III reverse transcriptase (Invitrogen) was subjected to previously described degenerate PCR conditions using the LPQG and YMDD primers (36). A 143-bp amplicon generated from the reaction was cloned into pCR2.1Topo vector (Invitrogen) and sequenced from M13F and M13R primers with an ABI 373A automated sequencer (Applied Biosystems, Inc., Foster City, CA) at the Biotechnology Resource Center at Cornell University. To obtain additional upstream retroviral sequences, degenerate primers to a third conserved region in *pro*, LVDTGA (5′-GTKTTIKTIGAYACIGGIKC-3′), and the YMDD primer were employed in RT-PCRs using previous reaction conditions. A 1,484-bp amplicon generated by this reaction was cloned into the pCR2.1Topo vector and DNA sequenced from M13F and M13R primers.

Cloning and sequencing of SSSV. A representative lambda library was constructed with salmon swim bladder sarcoma DNA partially digested with MboI and size selected on a 5 to 20% potassium acetate gradient; fragments of 15 to 20 kb were cloned and packaged using a lambda DASH II/BamHI vector kit (Stratagene, La Jolla, CA) using Gigapak II XL packaging extract according to the manufacturer's instructions. A library of 800,000 plaques was screened with a 143-bp *pol* sequence probe labeled with ^{32}P by random priming (Roche, Indianapolis, IN). Of the 150 positive clones obtained, twenty were plated and selected through two additional rounds of hybridization. Seventeen of these were analyzed by Southern blot; lambda DNAs were prepared from each clone, digested with PstI, separated on 1% agarose gels, and blotted onto nitrocellulose by standard procedures. All blots were hybridized with *pol* probes in 50% formamide buffer at 37°C for 24 h and washed in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 55°C. Fifteen positive clones were obtained and utilized for sequencing of the viral genome.

Subclones were PCR amplified from eleven overlapping lambda clones and used as a template for DNA sequencing by primer walking. Viral regions flanking the initially isolated *pol* sequence were PCR amplified with primers located on either end of the multiple cloning site of the lambda clones incorporating the forward T7 promoter (5′-CCGCGTAATACGACTCACTATAGGG-3′) and the reverse T3 promoter (5′-CCGCGAATTAACCCTCACTAAAGG-3′) were used in conjunction with forward (5′-TCAGTTCTCCCACAAGGGTACAGA-3′) or reverse (5′-ATGAGGACCACACCATCAGGGATT-3′) primers located within the *pol* region. PCRs were performed in a 100-μl volume containing 100 ng of lambda DNA, 20 mM Tris-HCl (pH 8.3), 2 mM $MgCl_2$, 50 mM KCl, 2.5% dimethyl sulfoxide, 200 μM (each) deoxynucleoside triphosphates, 10 pmol (each) of primer, and 2.5 U of *Taq* DNA polymerase (Invitrogen). The mixture was incubated 5 min at 94°C and then subjected to 35 cycles of 94°C for 30 s, 57°C for 20 s, and 72°C for 20 s, followed by a hold of 5 min at 72°C. PCR products were visualized on 1.2% agarose gels, cloned into pCR2.1Topo vector (Invitrogen), and sequenced with M13F and M13R primers. The SeqMan and Genequest programs from the Lasergene software package (DNASTAR, Inc., Madison WI) were utilized to assemble the sequence data into a contig containing the entire provirus and to search for open reading frames (ORFs). To further verify the presence of the complete provirus in salmon swim bladder sarcomas, regions encompassing *gag*, *pol*, *env*, and the long terminal repeats (LTRs) were PCR amplified, cloned, and sequenced from tumor DNA.

Southern and Northern blots. The proviral status of SSSV in tumors was analyzed by Southern blotting using the *pol* probe described above. Twenty-five milligrams of tumor tissue was homogenized into 200 μl of lysis buffer (0.1 M EDTA, 0.05 M Tris, pH 8) and then treated with 100 μg RNase A for 15 min at 22°C followed by 400 μg proteinase K and 0.1% sodium dodecyl sulfate for 2 h at 56°C. Genomic DNA from the resulting homogenate was extracted using the QIAGEN (Valencia, CA) miniprep blood kit protocol and reagents. One microgram of DNA from individual tumors from three separate Atlantic salmon and normal liver tissue from a captive Atlantic salmon from the Tunison Laboratory of Aquatic Sciences (U.S. Geological Society, Cortland, NY) was digested with PstI or KpnI restriction enzymes for 1 h at 37°C. The digested DNAs were separated in individual wells of a 1.2% agarose gel, blotted onto nitrocellulose, and probed with *pol* sequences using conditions described above.

Northern blot analysis of SSSV transcripts was accomplished by isolation of poly(A)$^+$ mRNA using a Poly(A)Purist kit (Ambion) from total RNA isolated from two individual salmon swim bladder sarcomas. Ten micrograms of poly(A)$^+$ RNA was separated in 1% formaldehyde gels, blotted onto nitrocellulose, and hybridized in formamide buffer at 37°C for 48 h to ^{32}P-labeled probes containing the U3 and R regions of the SSSV LTR. Blots were washed in 2× SSC at 55°C and exposed to film for 20 s. RNA molecular weight markers in sizes ranging from 0.16 kb to 1.77 kb and 0.24 kb to 9.5 kb (Invitrogen) were run in parallel to determine the size of the hybridizing signals.

Transcriptional mapping. Mapping of SSSV transcripts was accomplished by RT-PCR from DNase I-treated total RNA isolated from salmon swim bladder sarcomas. cDNA synthesis reactions employed 1 μg of total RNA, 10 pmol of antisense primers located in U3 (5′-TCCTGCTTCGGGCGTAGTCTCT-3′), *env* (5′-AGTCGACCCCATGGCGTTCTG-3′), or *pol* (5′-CGTTGTCGGATC TGATTTGTCGTG-3′), and Superscript III reverse transcriptase (Invitrogen). The cDNAs were amplified by PCR in a 100-μl volume containing 2 μl of cDNA, 20 mM Tris-HCl (pH 8.3), 2 mM $MgCl_2$, 50 mM KCl, 2.5% dimethyl sulfoxide, 200 μM (each) deoxynucleoside triphosphates, 2.5 U of *Taq* DNA polymerase (Invitrogen), and 10 pmol of forward primer located in the R region of the LTR (5′-GGCTCTCCCTTCTGACTCCCTCAG-3′) in combination with either the U3, *env*, or *pol* reverse primers. The mixture was incubated 5 min at 94°C and then subjected to 35 cycles of 94°C for 30 s, 57°C for 20 s, and 72°C for 20 s, followed by a hold of 5 min at 72°C. PCR products were obtained only from reactions containing the R forward and U3 or *env* reverse primer sets. Amplicons were gel purified from 1.2% agarose gels by Qiaex II (QIAGEN), cloned into the pCR2.1Topo vector, and sequenced with M13F and M13R primers. Splice donor and acceptor sites were determined by sequence comparison with the complete SSSV sequence contig.

Phylogenetic analysis. Predicted protein sequence comparisons against published sequences utilized the National Centers for Biotechnology Information Basic Local Alignment Search Tool (http://www.ncbi.nlm.nih.gov/BLAST/) (1). Each search utilized the BLOSUM 62 matrix, a minimum word size of 3, and no filtering. For the phylogenetic comparisons of reverse transcriptase amino acid sequences, the seven blocks of conserved sequences as defined by Xiong and Eickbush (52), consisting of 178 residues of reverse transcriptase, were located manually in the SSSV *pol* gene and aligned with the same sequences from other retroviruses as described previously (52). The 178 residues were treated as a single continuous sequence without gaps, and trees were generated using the neighbor-joining algorithm of the MEGA version 2.1 application (24). A total of 100 bootstrap replicates were conducted to confirm evolutionary relationships.

Nucleotide sequence accession number. The nucleotide sequence of the complete SSSV genome has been submitted to GenBank under the accession number DQ174103.

RESULTS

Identification of SSSV. Identification of retroviral sequences in Atlantic salmon swim bladder tumors was accomplished by RT-PCR with degenerate primers targeting conserved amino acid sequences in the retroviral RT gene (VLPQG and YMDD). By this technique, a 143-bp sequence with homology to previously described RT genes was amplified from tumor RNA. An additional upstream sequence was obtained by RT-PCR employing a degenerate primer based on a third conserved region of the viral protease (PR) (LVDTGA) in combination with the YMDD primer. The DNA sequence of the resulting 1,464-bp amplicon included the previously identified region of *pol* and an additional sequence with homology to previously sequenced retroviral *pol*.

To obtain the complete sequence of the SSSV provirus, a lambda library was constructed with salmon swim bladder sarcoma DNA partially digested with MboI restriction endonuclease. From a library of 800,000 plaques, approximately 150 plaques hybridized to a SSSV *pol* region probe. Given that

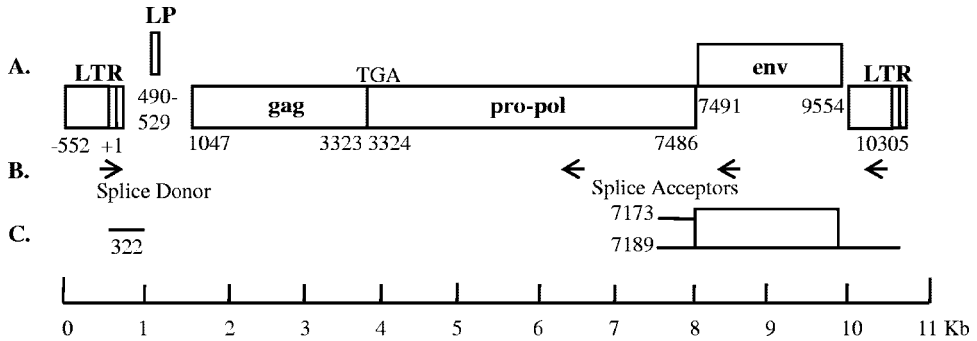

FIG. 1. (A) SSSV genome organization and splicing pattern. The numbers mark the boundaries of the ORFs and of the LTRs relative to the start of transcription (+1). The Gag-Pol polyprotein is predicted to be translated from a genomic RNA by suppression of a termination codon (TGA) at the junction of the *gag* and *pol* genes. LP, leader peptide. (B) Primers used for RT-PCR analysis of splice sites of SSSV. (C) Subgenomic viral transcripts. The positions of the major splice donor and acceptor sites relative to the start of transcription are indicated. The Env protein could be generated from either spliced subgenomic RNA.

approximately five library equivalents were screened (150,000 clones represent one library with an average insert of 15 to 20 kb) and 150 positive clones were obtained, we estimate that there are at least 30 proviral copies per tumor cell.

The complete 10,935-bp sequence of the salmon swim bladder sarcoma provirus was obtained from 11 overlapping lambda clones and sequenced to fivefold redundancy. Intact proviral sequences encompassing *gag*, *pol*, *env*, and the LTRs were further verified by PCR from tumor DNA and DNA sequencing. The organization of the SSSV provirus appears to be similar to that of gammaretroviruses with open reading frames that represent the *gag*, *pol*, and *env* genes flanked by LTR sequences at each end of the genome (Fig. 1A).

LTRs and leader. The LTRs comprise 713 bp of sequence at each end of the proviral genome. Each LTR is bounded by a short 3-bp inverted repeat containing the terminal sequences 5′-TG and 3′-CA found in all retroviruses. The SSSV LTRs are subdivided into three adjacent sequence blocks consisting of a 552-bp U3 region, an 83-bp R region, and a 78-bp U5 region (Fig. 1A). The site of transcriptional initiation (position +1), located by definition at the boundary of U3 and R, is predicted to initiate within the sequence GGGG. This is based upon a homology with WDSV of eight nucleotides encompassing the TATA box and a 75% homology over 20 bases surrounding the transcription initiation site. The putative SSSV core promoter localized within U3 contains a consensus TATA box (TATATAA) 23 bp upstream of the transcription start site as well as a GC box and two CCAAT boxes located 86 bp, 72 bp, and 102 bp upstream of the start point, respectively. In the 3′ LTR, the U3 region is preceded by a polypurine tract, GAA GAGAGGGA (position 9586 to 9597), likely to serve as a template for plus-strand synthesis during retroviral replication. In most retroviruses, the polyadenylation site of the viral transcript is preceded by a CA dinucleotide approximately 20 bp downstream of the polyadenylation signal. In SSSV, the CA dinucleotide is located only 5 bp downstream from a consensus polyadenylation signal, AATAAA (position 10295 to 10300), located within R. The 3′ end of U5 in the 5′ LTR is delineated by the presence of a tRNA-primer-binding site utilized during the plus-strand synthesis step of reverse transcription. SSSV contains a sequence in this region identical to the 3′ terminal 19 nucleotides of chum salmon (*Oncorhynchus keta*) initiator Met tRNA (12) (position 168 to 186), suggesting that this is the first retrovirus to utilize Met tRNA as a primer.

The leader region for SSSV, located proximal to the Gag start codon, is 1,045 nucleotides (nt) in length and contains two 51-nt direct repeats as well as three 20-nt and three 26-nt direct repeats. Within this region, SSSV contains a short leader region ORF initiating at a start codon 556 nt upstream of *gag* (position 490). Utilization of this start codon is predicted to encode a short 25-amino-acid (MQCNVTNCMIIWDTIHRC RQRPNEE) leader peptide with no homologous counterparts or conserved motifs (Fig. 1A).

Gag. The predicted start codon for *gag* is located at position 1047 on the viral RNA in a favorable sequence context for translation initiation (22), suggesting that its translation is dependent on leaky scanning of ribosomes (23) through the upstream leader peptide start codon. The *gag* ORF is predicted to encode a 758-amino-acid, 84-kDa polyprotein most similar to 533 amino acids in the N terminus of ZFERV Gag (25% identity) (Fig. 2) and to 179 amino acids in the capsid (CA) region of WEHV-1 (28% identity). Similar to snakehead retrovirus (SnRV) (14), the N terminus of SSSV Gag does not contain a consensus myristylation motif and may utilize an alternative mechanism for targeting to the plasma membrane. Similar to Mason-Pfizer monkey virus (MPMV) (53), murine leukemia virus (MuLV) (54), human T-cell leukemia virus type 1 (HTLV-1) (15, 48), and bovine leukemia virus (47), an L domain virion maturation sequence, PPPY, is located within a proline-rich region of Gag. Based on sequence comparison to WEHV-1 CA, we have identified a potential major homology region in Gag (position 2646), VVINGTPA PVKAKLRQTVGLC, which is quite divergent from the consensus established for other retroviruses (50). Similar to gammaretroviruses and epsilonretroviruses (46), the C-terminal region of Gag contains a single Cys-His box ($CX_2CX_4HX_4C$) at position 3096 which is utilized during viral genomic RNA packaging. Additional potential RNA encapsidation sequence motifs located downstream of the Cys-His box include glycine-arginine (GR)-rich regions previously described in the epsilonretroviruses (19, 25) and spumaviruses (27) and proline-glutamine-rich regions also present in epsilonretroviruses (19, 25).

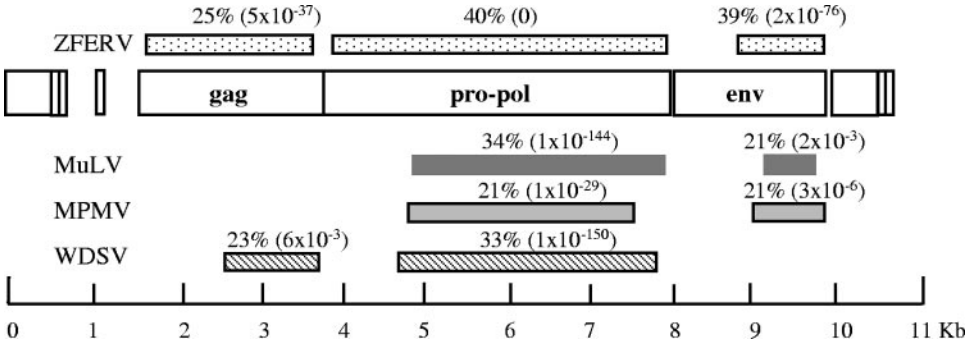

FIG. 2. Homology between SSSV and other retroviruses. Regions of homology between SSSV and ZFERV, MuLV, MPMV, and WDSV were determined by BLAST searches of putative SSSV Gag, Pol, and Env protein sequences. Regions of detectable protein identity between each virus and SSSV are delineated by overlapping bars. Percentages of identity between each protein in specified regions are noted above each bar with expected values in parentheses.

Pro and Pol. Expression of SSSV *pol* appears similar to that of the gammaretroviruses and epsilonretroviruses in its dependence on translational suppression of an in-frame stop codon located between *gag* and *pro-pol* (position 3321). By this process, a single large 238-kDa Gag-Pro-Pol precursor polyprotein 2,145 amino acids in length is predicted to be translated. Proteolytic processing of the polyprotein is necessary to generate the viral PR, RT, and integrase (IN).

Based on a distant sequence homology to gammaretroviruses, the predicted PR of SSSV is located in the first 199 amino acids of Pol and encoded in the same reading frame as RT and IN. The active-site aspartate residue of PR is located within the conserved LVDT sequence motif at position 3420 and a conserved GRDXL motif is located at position 3612. The amino acid sequence of SSSV PR is most closely related to those of ZFERV (34% identity), MuLV (32% identity), and feline leukemia virus (28% identity).

Homology to the RT and RNase H domains of retroviruses spans the central region of SSSV Pol. This region is most homologous to that of ZFERV (40% identity) and to a lesser extent of MuLV (34% identity) and WDSV (33% identity). The conserved polymerase sequences LPQG and YVDD are located at positions 4773 and 4871 of SSSV *pol*, respectively. A region downstream corresponding to the predicted SSSV IN contains a conserved HHCC N-terminal zinc finger domain enzyme motif (position 6476) and a DD35E central catalytic domain (position 6686). This region has the highest identity to that of ZFERV (42% identity), reticuloendotheliosis virus (34% identity), and of WDSV (32% identity).

A 21-nucleotide purine-rich region located in *pol* at position 5376 has homology to central polypurine tract (cPPT) sequences similarly identified in lentiviruses (3, 5, 20, 42, 43, 49) and spumaviruses (39, 45) (Table 1). In human immunodeficiency virus type 1 (HIV-1) and other lentiviruses (28, 32), utilization of the cPPT as an internal priming site during plus-strand synthesis has been demonstrated to generate a central DNA flap that enhances infection efficiency in nondividing cells (55). The proper displacement of the central DNA flap is dependent on adjacent central termination sequences (CTS) that terminate plus-strand synthesis initiated from the 3′ LTR (6, 26). Distinguishing features of the CTS are multiple A-tracts, which may induce localized distortion of duplex DNA (43, 49). In SSSV, we have located two potential CTS regions, 79 nt and 90 nt downstream of the cPPT.

Env. The start codon for *env* is located at position 7491 and appears dependent on a singly spliced viral transcript for expression. Expression of *env* is predicted to encode a 687-amino-acid protein sharing highest identity to ZFERV (39%) and betaretroviruses such as MPMV (21%) in the central and C-terminal regions. SSSV Env does not show any significant sequence similarity to those of epsilonretroviruses. A 23-amino-acid hydrophobic region located 29 amino acids from the N terminus of Env is consistent with a signal sequence utilized for translocation into the endoplasmic reticulum. N-linked glycosylation of Env is predicted at 14 sites that fit the consensus sequence Asn-X-Ser or Asn-X-Thr (where X is any amino acid except proline) (16). In addition, proteolytic cleavage of Env to surface (SU) and transmembrane (TM) is predicted at a furin recognition sequence (RSKR) [consensus Arg-X-(Lys/Arg)-Arg] (44) located at position 8862. A hydrophobic fusion peptide region, typically found downstream of the SU/TM cleavage site in other retroviral TM proteins, could not easily be identified in SSSV. The C-terminal region of Env contains a conserved immunosuppressive domain (position 9126) followed by a cystine-rich sequence, predicted to form disulfide bonds. A hydrophobic region spanning positions 9354 to 9440 represents

TABLE 1. Alignment of predicted SSSV cPPT with cPPT regions identified in complex retroviruses

Retrovirus	Predicted cPPT DNA sequence[a]
SSSV	AAACAAGAGGGGGAGAGAAGGG
HIV-1	TTTTAAAAGAAAAGGGGGGATTGGG
SIVmac239	TTTTAAAAGAAGGGGAGGAATAGG
FIV	TTTTAAAAGAAGAGGTAGGATAGGA
Visna virus	TTTTAAAAGAAGAGGTAGGATAGGA
EIAV	TTGTAACAAAGGGAGGGAAAGTATG
BIV	TTGTAACAAAGGGAGGGAAAGTATG
HFV	GTCCAGGAGAGGGTGGCTAGG

[a] cPPT regions from viral sequences of HIV-1 (6), simian immunodeficiency virus strain mac239 (SIVmac239) (30), feline immunodeficiency virus (FIV) (51), visna virus (3), equine infectious anemia virus (EIAV) (45), bovine immunodeficiency virus (BIV) (34), and human foamy virus (HFV) (46) were derived from experimental data.

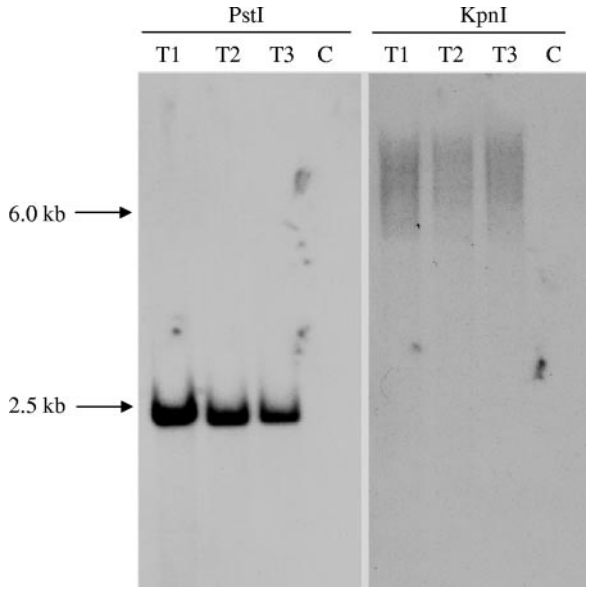

FIG. 3. SSSV integration pattern in three salmon swim bladder sarcomas (T1 to T3) and control liver tissue (C). Genomic DNA cut with either PstI or KpnI and probed for SSSV *pol* sequences by Southern blotting. PstI sites are located upstream and downstream of SSSV *pol* and thus create a 2.5-kb fragment after PstI digestion. Enzyme digestion with KpnI produces a broad range of fragments greater than 6 kb due to the presence of a single KpnI site located upstream of *pol* and random sites in the Atlantic salmon genome.

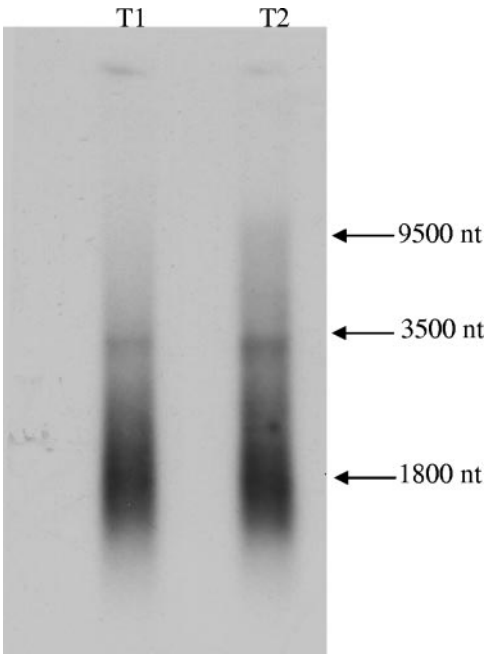

FIG. 4. Northern blot analysis of SSSV expression in swim bladder sarcoma tumors. Poly(A)$^+$ RNA isolated from two individual tumors (T1 and T2) were probed with 552-bp DNA sequences from the U3 and R regions of SSSV LTR. The sizes of transcripts were determined by comparison to an RNA ladder marker run in parallel.

the potential membrane-spanning region of the protein that would result in a 39-amino-acid cytoplasmic domain.

Proviral DNA status in swim bladder sarcomas and normal tissues. The integration status of SSSV DNA in salmon swim bladder sarcoma tumors and normal salmon tissue was assessed by Southern blotting with a SSSV *pol* probe (Fig. 3). Genomic DNA from three tumor samples digested with PstI produced hybridization signals that correlated to the predicted 2.5 kb of sequence between PstI restriction sites within SSSV *pol* (Fig. 3). In contrast, control genomic DNA isolated from a captive Atlantic salmon was negative, indicating that SSSV *pol* sequences are not endogenous within the Atlantic salmon genome. Southern blot analysis of tumor DNA samples digested with KpnI, which cleaves once upstream of *pol*, yielded a broad smear beginning at about 6 kb in all three tumor samples, but not in control DNA (Fig. 3). This pattern is consistent with a high copy number and polyclonal integration of SSSV in all three swim bladder sarcomas.

TABLE 2. Sequences and locations of SSSV splice donor and acceptor sites

Splice donor or acceptor sequence	Position
Donor	
NAG/GTRAGT	Consensus
AAG/GTAAGC	322
Acceptor	
YYTTYYYYYYNCAG/G	Consensus
CGAATTATCTGCAG/CTCT	7173
CTCTCTCCATACAG/ATAC	7187

Expression of SSSV transcripts in salmon swim bladder sarcomas. To map spliced SSSV transcripts and associated splice donor and acceptor sites, cDNAs were synthesized from tumor RNA using a forward primer based on the R sequence and reverse primers based on U3, *env*, or *pol* sequences in RT-PCRs (Fig. 1B). Single amplicons were obtained with the LTR and *env* primers, but not with *pol* primers, suggesting the primary viral transcript in swim bladder sarcomas is singly spliced (Fig. 1C). At the splice junctions, joined sequences contain a single splice donor site upstream of *gag* (position 322) and two overlapping splice acceptor sites at positions 7173 and 7187 of the virus (Fig. 1C). While the 5′ SSSV splice donor is in agreement with the consensus sequence (40) (Table 2), the two 3′ splice acceptor sites deviate from consensus sequence. Deviation from canonical splice acceptor sequences has been found in other retroviruses, such as bovine syncytial virus (35) and the walleye retroviruses (19, 25, 37).

To detect SSSV expression in tumors, poly(A)$^+$ RNA was isolated from salmon swim bladder sarcomas and probed for sequences in the R and U5 region of the 3′ LTR by Northern blot. A smear made up of high-molecular-weight RNAs initiating at approximately 10,000 nt likely corresponds to full-length genomic viral transcripts (Fig. 4). A defined band slightly less than 3,500 nt is suggestive of singly spliced viral transcripts predicted to be either 3,454 nt and 3,444 nt, depending on selection of splice acceptor sites. A low-molecular-weight signal at 1,800 nt indicates that RNA was slightly degraded in the tumor samples analyzed.

Phylogenetic analysis. To determine the relationship between SSSV and prototypic members of established retrovirus genera, an unrooted neighbor-joining phylogenetic tree was

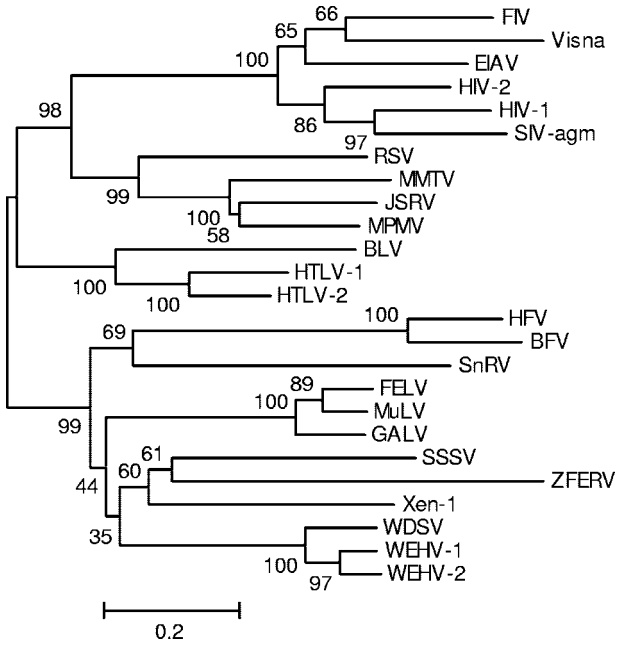

FIG. 5. Unrooted phylogenetic tree of representative retroviruses based on an amino acid alignment of seven conserved domains in reverse transcriptase (52) totaling 178 residues. Retroviruses are designated as follows: MPMV, JSRV, MMTV (mouse mammary tumor virus), RSV (Rous sarcoma virus), EIAV (equine infectious anemia virus), FIV (feline immunodeficiency virus), Visna (visna virus), HIV-1, HIV-2, SIV-agm (simian immunodeficiency virus-agm), HTLV-1, HTLV-2, BLV (bovine leukemia virus), SSSV, SnRV, BFV (bovine foamy virus), HFV (human foamy virus), FeLV (feline leukemia virus), MuLV, GALV (gibbon ape leukemia virus), WDSV, WEHV-1, WEHV-2, ZFERV, and Xen-1. Phylogenetic and molecular evolutionary analyses were conducted using the neighbor-joining methods of MEGA version 2.1 software (24). Bootstrap values displayed at each branch point were determined from 100 replicates. Horizontal branch lengths are proportional to the degree of amino acid substitutions per site.

constructed using alignments of seven conserved regions of the retroviral Pol (52). Sequences from completely sequenced retroviruses from lower vertebrates were included in the alignment. The resulting tree places SSSV closest to the *Epsilonretrovirus* and *Gammaretrovirus* genera, but in a distinct branch including ZFERV and *Xenopus* endogenous retrovirus-1 (Xen-1) (21) (Fig. 5). In this tree, SnRV groups separately from the other piscine retroviruses in a distinct branch most closely related to spumaviruses (Fig. 5). While the tree illustrates that SSSV has diverged from the epsilonretroviruses, the low bootstrap value for this group indicates that the alignment cannot unambiguously resolve the branching order of SSSV and the *Epsilonretrovirus* and *Gammaretrovirus* genera.

DISCUSSION

An exogenous piscine retrovirus, SSSV, has been isolated from an outbreak of leiomyosarcoma in the swim bladders of commercially grown Atlantic salmon. The structural and transcriptional profile of SSSV indicates that it is a simple retrovirus; there is, at most, only one potential coding region in the leader in addition to *gag*, *pol*, and *env*. Consequently, SSSV appears different from exogenous piscine retroviruses of the *Epsilonretrovirus* genus which express two ORFs downstream of *env* (18). Although structurally similar to simple retroviruses, SSSV is distinguished from them in that no related endogenous counterparts have been identified in the Atlantic salmon genome. Simple exogenous retroviruses of the *Alpharetrovirus*, *Betaretrovirus*, and *Gammaretrovirus* genera share multiple copies of related endogenous sequences fixed in host genomes by recurrent integration events in germ line DNA (13). The lack of endogenous copies of SSSV in the Atlantic salmon genome could be explained by a recent zoonotic transmission of SSSV into Atlantic salmon.

The high copy numbers of proviruses (greater than 30 copies per cell) in salmon swim bladder sarcomas are also rare for a simple retrovirus. High proviral copy numbers have previously been reported in complex retrovirus infection of several cell types. Infection of human erythroleukemia cells by the spumaviruses human and simian foamy virus have been reported at levels of 20 provirus copies per cell (31). Additionally, high copies of unintegrated viral DNA have been documented in WDSV-associated tumors (10 to 50 copies per cell) (29) and HIV-1 infection of human lymphocytes (20 to 400 copies per cell) (2). The accumulation of large numbers of proviruses suggests that SSSV is expressed at high levels in salmon swim bladder sarcomas leading to multiple infections of tumor cells.

A feature common to most sequenced piscine retroviruses (19, 25, 41) and Xen-1 (21) are long leader sequences, ranging from 750 nt in Xen-1 to 1,195 nt in ZFERV, containing several direct repeats. Additionally, the leader regions of these retroviruses contain short ORFs that initiate upstream of the *gag* start codon. The function of leader peptides remains largely unknown in retroviruses. In WDSV, a 120-amino-acid leader ORF, OrfC, has been demonstrated to localize to mitochondria and increase apoptosis in expressing cells (33). The leader peptide of SnRV is predicted to initiate translation and fuse to the *env* ORF by use of a splice donor located within the coding region of the leader peptide and a splice acceptor in the 3′ region of *pol* (14). Rous sarcoma virus contains three ORFS in the leader region that encode peptides 7, 16, and 9 amino acids in length demonstrated to affect translational initiation from the Gag start codon (8). By BLAST analysis, the predicted 25-amino-acid leader peptide encoded by SSSV does not contain convincing homology to any known protein or functional domain. Experimental evidence will be needed to determine the extent and role of LP production in SSSV-infected cells.

The identification of SSSV sequences associated with an outbreak of the swim bladder sarcoma suggests a role for the retrovirus in pathogenesis of the disease. The high copy number and polyclonal pattern of integration of SSSV within tumors seemingly excludes the possibility of insertional activation of a proto-oncogene as a mechanism of tumorigenesis. However, high copy numbers of proviruses in tumors acquired after the tumor initiation event may mask an insertional activation event arising within tumors.

It is feasible that the oncogenic potential of SSSV may arise as the result of the expression of a viral gene product. This mechanism of oncogenesis has been proposed for the Jaagsiekte sheep retrovirus (JSRV) in the induction of ovine pulmonary adenocarcinoma (17, 51) and in Friend spleen focus-forming virus-associated erythroid hyperplasia in mice (38). In

cell culture and mouse models, the Env protein of JSRV has been demonstrated to induce transformation through a YXXM motif (a putative binding site for phosphatidyl-inositol 3-kinase) located in the cytoplasmic domain of TM in combination with uncharacterized regions of SU (17, 51). Additionally, the presence of common integration sites within JSRV-associated tumors suggests that insertional mutagenesis may act in concert with the viral protein in tumor development (7). Although a related YXXM signaling motif has not been identified in the predicted cytoplasmic domain of SSSV TM, further experimental analysis is needed to establish the oncogenic potential of SSSV proteins.

A multifactorial mechanism for the development of salmon swim bladder sarcoma must also be considered. Preliminary laboratory transmission experiments using filtered SSSV tumor homogenates have been successful at transmitting SSSV infection to naive Atlantic salmon; however, tumors were not observed (P. Bowser, personal communication). Thus, the presence of other infectious agents as well as host or environmental cofactors cannot be excluded. It is intriguing that the swim bladder appears to be a target tissue for SSSV infection and SSSV-associated tumors. It is unclear if pathology in the swim bladder is a result of viral preference for infection of that tissue type or through the activation of tissue-specific cofactors.

Phylogenetic analysis based on conserved regions of Pol places SSSV in a novel group positioned between the *Gammaretrovirus* and *Epsilonretrovirus* genera. This model is supported by high bootstrap values between the spumaviruses and the gammaretroviruses and epsilonretroviruses and homology between portions of SSSV Gag and Pol to similar proteins of gammaretroviruses and epsilonretroviruses. Structural features such as a single Cys-His box in Gag and predicted amber suppression for translation of the Gag-Pol polyprotein are also characteristic of gammaretroviruses and epsilonretroviruses. Given the low bootstrap values between the SSSV and the gammaretroviruses and epsilonretroviruses, it is not possible at this point to accurately place the SSSV in relationship with either of these genera. However, the homology to betaretroviruses in Env, possible utilization of a cPPT, and usage of a Met-tRNA primer for first-strand synthesis indicate that SSSV possesses features distinct from viruses of the genus *Gammaretrovirus* or *Epsilonretrovirus*. Sequenced endogenous retroviruses, ZFERV and Xen-1 also align with the SSSV. While ZFERV has a high degree of homology with SSSV extending over large portions of Gag, Pol, and Env, Xen-1 appears significantly divergent outside of the RT domain. Although ZFERV is an endogenous retrovirus, all ORFs are intact and the provirus appears to have a polymorphic integration pattern within the zebrafish genome (41). Thus, ZFERV appears to represent a recent acquisition within the zebrafish genome and, like other exogenous simple retroviruses, may possess an exogenous counterpart. Further discovery of similar retroviruses will be needed to validate the existence and extent of this new group of retroviruses.

Although the outcome of SSSV infection in wild populations has not been thoroughly investigated, the lethargic presentation of infected salmon and manifestation of swim bladder tumors suggest that this virus may have a negative impact on the natural salmon population (4). It is likely that conditions such as increased viral transmission and stress related to rearing of salmon in captivity may play a role in the etiology of the disease in these populations. Understanding the factors involved in outbreaks of salmon swim bladder sarcoma will require a greater understanding of the oncogenic potential of SSSV and the expression pattern of the virus in infected fish. Additional molecular characterization of SSSV will assist in the development of diagnostic tools to determine the extent of SSSV infection in Atlantic salmon populations and prevent future outbreaks of the disease.

ACKNOWLEDGMENTS

We thank Brian Lazzaro for assistance with phylogenetic analyses and Alan Eaglesham for editorial review.

This research was supported in part by USDA grants 99-35204-7485 and 02-35204-12777 to J.W.C. This work was also supported in part by the National Oceanic and Atmospheric Administration award no. NA86RG0056 to the Research Foundation of State University of New York for New York Sea Grant to P.R.B. T.A.P. was supported by NIH training grant 5T32CA09682.

The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies.

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Bell, P., L. J. Montaner, and G. G. Maul.** 2001. Accumulation and intranuclear distribution of unintegrated human immunodeficiency virus type 1 DNA. J. Virol. **75:**7683–7691.
3. **Blum, H. E., J. D. Harris, P. Ventura, D. Walker, K. Staskus, E. Retzel, and A. T. Haase.** 1985. Synthesis in cell culture of the gapped linear duplex DNA of the slow virus visna. Virology **142:**270–277.
4. **Bowser, P. R., J. W. Casey, R. N. Casey, S. L. Quackenbush, L. Lofton, J. A. Coll, and R. C. Cipriano.** Submitted for publication.
5. **Charneau, P., and F. Clavel.** 1991. A single-stranded gap in human immunodeficiency virus unintegrated linear DNA defined by a central copy of the polypurine tract. J. Virol. **65:**2415–2421.
6. **Charneau, P., G. Mirambeau, P. Roux, S. Paulous, H. Buc, and F. Clavel.** 1994. HIV-1 reverse transcription. A termination step at the center of the genome. J. Mol. Biol. **241:**651–662.
7. **Cousens, C., J. V. Bishop, A. W. Philbey, C. A. Gill, M. Palmarini, J. O. Carlson, J. C. DeMartini, and J. M. Sharp.** 2004. Analysis of integration sites of Jaagsiekte sheep retrovirus in ovine pulmonary adenocarcinoma. J. Virol. **78:**8506–8512.
8. **Donze, O., P. Damay, and P.-F. Spahr.** 1995. The first and third uORFs in RSV leader RNA are efficiently translated: implications for translational regulation and viral RNA packaging. Nucleic Acids Res. **23:**861–868.
9. **Duncan, I. B.** 1978. Evidence for an oncovirus in the swim bladder fibrosarcoma of Atlantic salmon *Salmo salar* L. J. Fish Dis. **1:**127–131.
10. **Fan, H.** 1994. Retroviruses and their role in cancer, p. 313–362. *In* J. A. Levy (ed.), The Retroviridae, vol. 3. Plenum Press, New York, New York.
11. **Getchell, R. G., J. W. Casey, and P. R. Bowser.** 1998. Seasonal occurrence of virally induced skin tumors in wild fish. J. Aquat. Anim. Health **10:**191–201.
12. **Gillum, A. M., N. Urquhart, M. Smith, and U. L. RajBhandary.** 1975. Nucleotide sequence of salmon testes and salmon liver cytoplasmic initiator tRNA. Cell **6:**395–405.
13. **Gifford, R., and M. Tristem.** 2003. The evolution, distribution, and diversity of endogenous retroviruses. Virus Genes **26:**291–315.
14. **Hart, D., G. N. Frerichs, A. Rambaut, and D. E. Onions.** 1996. Complete nucleotide sequence and transcriptional analysis of the snakehead retrovirus. J. Virol. **70:**3606–3616.
15. **Heidecker, G., P. A. Lloyd, K. Fox, K. Nagashima, and D. Derse.** 2004. Late assembly motifs of human T-cell leukemia virus type 1 and their relative roles in particle release. J. Virol. **78:**6636–6648.
16. **Helenius, A., and M. Aebi.** 2004. Roles of N-linked glycans in the endoplasmic reticulum. Annu. Rev. Biochem. **73:**1019–1049.
17. **Hofacre, A., and H. Fan.** 2004. Multiple domains of the Jaagsiekte sheep retrovirus envelope protein are required for transformation of rodent fibroblasts. J. Virol. **78:**10479–10489.
18. **Holzschu, D., L. A. Lapierre, and M. D. Lairmore.** 2003. Comparative pathogenesis of epsilonretroviruses. J. Virol. **77:**12385–12391.
19. **Holzschu, D. L., D. Martineau, S. K. Fodor, V. M. Vogt, P. R. Bowser, and J. W. Casey.** 1995. Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. J. Virol. **69:**5320–5331.
20. **Hungnes, O., E. Tjotta, B. Grinde.** 1991. The plus strand is discontinuous in a subpopulation of unintegrated HIV-1 DNA. Arch. Virol. **116:**133–141.
21. **Kambol, R., P. Kabat, and M. Tristem.** 2003. Complete nucleotide sequence

of an endogenous retrovirus from the amphibian, *Xenopus laevis*. Virology **311:**1–6.
22. **Kozak, M.** 1987. An analysis of 5′-noncoding regions from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15:**8125–8148.
23. **Kozak, M.** 2002. Pushing the limits of the scanning mechanism for initiation of translation. Gene **299:**1–34.
24. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
25. **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.
26. **Lavigne, M., P. Roux, H. Buc, and F. Schaeffer.** 1997. DNA curvature controls termination of plus strand DNA synthesis at the centre of HIV-1 genome. J. Mol. Biol. **266:**507–524.
27. **Lochelt, M., and R. M. Flugel.** 1995. The molecular biology of human and primate spuma retroviruses, p. 239–292. *In* J. A. Levy (ed.), The Retroviridae, vol. 4. Plenum Press, New York, New York.
28. **Mangeot, P. E., D. Negre, B. Dubois, A. J. Winter, P. Leissner, M. Mehtali, D. Kaiserlian, F. L. Cosset, and J. L. Darlix.** 2000. Development of minimal lentivirus vectors derived from simian immunodeficiency virus (SIVmac251) and their use for gene transfer into human dendritic cells. J. Virol. **74:**8307–8315.
29. **Martineau, D., R. Renshaw, J. R. Williams, J. W. Casey, and P. R. Bowser.** 1991. A large unintegrated retrovirus DNA species present in a dermal tumor of walleye *Stizostedion vitreum*. Dis. Aquat. Org. **10:**153–158.
30. **McKnight, I. J.** 1978. Sarcoma of the swim bladder of Atlantic salmon (*Salmo salar* L.). Aquaculture **13:**55–60.
31. **Meiering, C. D., K. E. Comstock, and M. L. Linial.** 2000. Multiple integrations of human foamy virus in persistently infected human erythroleukemia cells. J. Virol. **74:**1718–1726.
32. **Molina, R. P., M. Matukonis, B. Paszkiet, J. Zhang, M. Kaleko, and T. Luo.** 2002. Mapping of the bovine immunodeficiency virus packaging signal and RRE and incorporation into a minimal gene transfer vector. Virology **304:**10–23.
33. **Nudson, W. A., J. Rovnak, M. Buechner, and S. L. Quackenbush.** 2003. Walleye dermal sarcoma virus Orf C is targeted to the mitochondria. J. Gen. Virol. **84:**375–381.
34. **Quackenbush, S. L., J. Rovnak, R. N. Casey, T. A. Paul, P. R. Bowser, C. Sutton, and J. W. Casey.** 2001. Genetic relationship of tumor-associated piscine retroviruses. Mar. Biotechnol. **3:**S88–S99.
35. **Renshaw, R. W., and J. W. Casey.** 1994. Transcriptional mapping of the 3′ end of the bovine syncytial virus genome. J. Virol. **68:**1021–1028.
36. **Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff.** 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res. **26:**1628–1635.
37. **Rovnak, J., J. W. Casey, and S. L. Quackenbush.** 2001. Intracellular targeting of walleye dermal sarcoma virus Orf A (rv-cyclin). Virology **280:**31–40.
38. **Ruscetti, S. K.** 1999. Deregulation of erythropoiesis by the Friend spleen focus-forming virus. Int. J. Biochem. Cell Biol. **31:**1089–1109.
39. **Schweizer M., R. Renne, and D. Neumann-Haefelin.** 1989. Structural analysis of proviral DNA in simian foamy virus (LK-3)-infected cells. Arch. Virol. **109:**103–114.
40. **Senapathy, P. P., M. B. Shapiro, and N. L. Harris.** 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol. **183:**252–278.
41. **Shen, C. H., and L. A. Steiner.** 2004. Genome structure and thymic expression of an endogenous retrovirus in zebrafish. J. Virol. **78:**899–911.
42. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **43:**369–382.
43. **Stetor, S. R., J. W. Rausch, M. J. Guo, J. P. Burnham, L. R. Boone, M. J. Waring, and S. F. Le Grice.** 1999. Characterization of (+) strand initiation and termination sequences located at the center of the equine infectious anemia virus genome. Biochemistry **38:**3656–3667.
44. **Thomas, G.** 2002. Furin at the cutting edge: from protein traffic to embryogenesis and disease. Nat. Rev. Mol. Cell Biol. **3:**753–766.
45. **Tobaly-Tapiero, J., J. J. Kupiec, M. Santillana-Hayat, M. Canivet, J. Peries, and R. Emanoil-Ravier.** 1991. Further characterization of the gapped DNA intermediates of human spumavirus: evidence for a dual initiation of plus-strand DNA synthesis. J. Gen. Virol. **72:**605–608.
46. **Vogt, V. M.** 1997. Retroviral virions and genomes, p. 27–69. *In* J. M. Coffin, S. H. Hughes, and H. E. Varmus (ed.), Retroviruses. CSHL Press, New York, New York.
47. **Wang, H., K. M. Norris, and L. M. Mansky.** 2002. Analysis of bovine leukemia virus Gag membrane targeting and late domain function. J. Virol. **76:**8485–8493.
48. **Wang, H., N. J. Machesky, and L. M. Mansky.** 2004. Both the PPPY and PTAP motifs are involved in human T-cell leukemia virus type 1 particle release. J. Virol. **78:**1503–1512.
49. **Whitwam, T., M. Peretz, and E. Poeschla.** 2001. Identification of a central DNA flap in feline immunodeficiency virus. J. Virol. **75:**9407–9414.
50. **Wills, J. W., and R. C. Craven.** 1991. Form, function, and use of retroviral Gag proteins. AIDS **5:**639–654.
51. **Wootton, S. K., C. L. Halbert, and A. D. Miller.** 2005. Sheep retrovirus structural protein induces lung tumours. Nature **434:**904–907.
52. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.
53. **Yasuda, J., and E. Hunter.** 1998. A proline-rich motif (PPPY) in the Gag polyprotein of Mason-Pfizer monkey virus plays a maturation-independent role in virion release. J. Virol. **72:**4095–4103.
54. **Yuan, B., X. Li, and S. P. Goff.** 1999. Mutations altering the moloney murine leukemia virus p12 Gag protein affect virion production and early events of the virus life cycle. EMBO J. **18:**4700–4710.
55. **Zennou, V., C. Petit, D. Guetard, U. Nerhbass, L. Montagnier, and P. Charneau.** 2000. HIV-1 genome nuclear import is mediated by a central DNA flap. Cell **101:**173–185.