Molecular Characterization of a Novel Gammaretrovirus in Killer Whales (*Orcinus orca*)[▽]

Sarah A. LaMere,¹ Judy A. St. Leger,² Mark D. Schrenzel,³ Simon J. Anthony,³
Bruce A. Rideout,³ and Daniel R. Salomon^{1*}

*Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037¹;
SeaWorld San Diego, San Diego, California 92109²; and Institute for Conservation Research, San Diego Zoo,
15600 San Pasqual Valley Road, Escondido, California 92027³*

Received 1 July 2009/Accepted 30 September 2009

There are currently no published data documenting the presence of retroviruses in cetaceans, though the occurrences of cancers and immunodeficiency states suggest the potential. We examined tissues from adult killer whales and detected a novel gammaretrovirus by degenerate PCR. Reverse transcription-PCR also demonstrated tissue and serum expression of retroviral mRNA. The full-length sequence of the provirus was obtained by PCR, and a TaqMan-based copy number assay did not demonstrate evidence of productive infection. PCR on blood samples from 11 healthy captive killer whales and tissues from 3 free-ranging animals detected the proviral DNA in all tissues examined from all animals. A survey of multiple cetacean species by PCR for *gag*, *pol*, and *env* sequences showed homologs of this virus in the DNA of eight species of delphinids, pygmy and dwarf sperm whales, and harbor porpoises, but not in beluga or fin whales. Analysis of the bottlenose dolphin genome revealed two full-length proviral sequences with 97.4% and 96.9% nucleotide identity to the killer whale gammaretrovirus. The results of single-cell PCR on killer whale sperm and Southern blotting are also consistent with the conclusion that the provirus is endogenous. We suggest that this gammaretrovirus entered the delphinoid ancestor's genome before the divergence of modern dolphins or that an exogenous variant existed following divergence that was ultimately endogenized. However, the transcriptional activity demonstrated in tissues and the nearly intact viral genome suggest a more recent integration into the killer whale genome, favoring the latter hypothesis. The proposed name for this retrovirus is killer whale endogenous retrovirus.

Retroviruses have been linked to a variety of diseases in a wide array of different vertebrate species. Most commonly, they have been implicated in neoplastic and immunodeficiency disorders, although they have also been demonstrated as etiologic agents in neurologic and respiratory diseases (8, 15, 16, 18, 28–30, 35, 38, 49). Exogenous retroviruses, which are not present in the germ line and are most often horizontally transmitted, are more frequently associated with disease. Endogenous retroviruses, which are incorporated into germ line DNA and may be transmitted vertically or horizontally, have also been linked to disease, including neoplastic disorders of mice and koalas (15, 46, 49–51, 58). However, in most cases (particularly autoimmune disorders of humans), it has been difficult to associate them with specific diseases because of the confounding presence of the virus in all individuals of a given host species (11, 12, 25, 36, 41, 57). In addition, while a correlation between endogenous retroviral expression and disease may be demonstrated in some instances, genes are often aberrantly expressed in disease states, and expression of an endogenous provirus might be an effect of the disease rather than a cause (8, 56, 58).

Most endogenous retroviruses are not associated with disease, though in many instances it is assumed that there was a

disease association prior to the virus entering the germ line (endogenization) (8, 56–58). Following endogenization, evolution tends to select for defective retroviruses, since maintaining a pathogenic virus in the germ line of a species would be actively selected against. Thus, most genomic loci of endogenous retroviruses accumulate mutations gradually rendering them replication incompetent. In situations where the virus has remained replication competent, the host species develops other methods of adaptation that prevent productive infection or disease (8).

Unfortunately, the distinction between endogenous and exogenous retroviruses is not always straightforward, and the presence of related endogenous retroviruses has been suggested to affect the course of exogenous retroviral infection. One example of this phenomenon is feline leukemia virus (FeLV), where full-length, functional genomes of endogenous FeLVs are found in domestic cats, and the difference between endogenous FeLV and exogenous FeLV is only in the U3 sequences of their long terminal repeats (LTRs) (6, 39, 47, 48). It has been reported that endogenous and exogenous FeLV variants will recombine to produce new exogenous variants (44). In addition, it has been suggested that immune tolerance can develop to exogenous retroviruses that look very similar to endogenous antigens recognized as “self,” resulting in immune evasion. Conversely, endogenous retroviruses have also been reported to block receptors, thus preventing cell entry by their exogenous counterparts (8, 17, 48, 58).

The possibility of retroviral infection has been raised in some cases of marine mammal neoplasia, such as immunoblas-

* Corresponding author. Mailing address: Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037. Phone: (858) 784-9381. Fax: (858) 784-2121. E-mail: dsalomon@scripps.edu.

[▽] Published ahead of print on 7 October 2009.

tic malignant lymphoma in dolphins, several neoplastic diseases of beluga whales, and malignant leukemia/lymphoma found in harbor seals (3, 31, 60). In particular, neoplasia is frequently reported in delphinids, including the killer whale (3, 31, 60). However, neoplastic diseases of marine mammals are generally hypothesized to be related to pollution of waterways and estuaries. For example, beluga whales in the St. Lawrence Estuary in Canada have a high incidence of neoplasia (37%), and large amounts of chemical carcinogens have been documented in belugas from the gulf of St. Lawrence versus those in the arctic (3, 31, 60).

While environmental exposures, genetic mutations, and retroviral infections are known causes of neoplasia, the role of retroviral infection in neoplasia of marine mammals cannot be determined until such retroviruses are identified and better understood. It is also probable that the cause of neoplasia in marine mammals is multifactorial, even if a retroviral component exists. For example, it was recently discovered that an exogenous murine leukemia-related gammaretrovirus was associated with some cases of prostate cancer in humans, but infection with this virus was correlated with a mutation in a specific host gene linked to the cell-mediated immune response (55). As environmental contamination and human encroachment become increasingly controversial issues with regard to the survival and health of marine mammal species, it is important to understand potential infectious agents affecting marine mammals and their interplay with host genetics and environment.

Only one retrovirus has ever been identified in a marine mammal: an exogenous spumavirus that was isolated from a California sea lion (*Zalophus californianus*) (19). No retroviruses have been established in any cetacean species, endogenous or exogenous, although there is one report of a polymerase sequence in a Risso's dolphin (*Grampus griseus*) that was presumably derived from a betaretrovirus (14). In the present article, we provide the first documentation of a full-length endogenous retrovirus in cetaceans, including molecular characterization and phylogenetic analysis. The proposed name for this retrovirus is killer whale endogenous retrovirus (KWERV), though analysis of the different cetacean species related to the killer whale suggests the presence of this retroviral sequence in all delphinids.

MATERIALS AND METHODS

Nucleic acid isolation. All samples used for this work were collected opportunistically. Blood samples were collected from 11 captive killer whales. Tissue samples were collected from two captive killer whales and three deceased free-ranging killer whales. All tissue samples collected were stored at -80°C and transported on dry ice. Blood samples were collected into tubes containing EDTA and either processed immediately or frozen at -80°C . Genomic DNA was prepared using the Qiagen DNeasy blood and tissue kit. Blood samples for RNA extraction were transported in PaxGene tubes and processed per the manufacturer's protocol or tubes containing EDTA on ice and extracted using Trizol (Life Technologies). Plasma RNA was extracted using the Qiagen viral RNA kit. For each sample, 1 μg of RNA was DNase treated with Invitrogen reverse transcription-grade DNase I for 15 min at room temperature. DNase was disabled after 15 min with EDTA (2.5 mM) at 65°C for 10 min. cDNA was prepared with Invitrogen Superscript III or with Bio-Rad iScript according to the manufacturer's instructions. For each cDNA sample, a control without reverse transcriptase was performed, and a rRNA PCR was done to rule out genomic DNA contamination.

Initial gammaretrovirus amplification and cloning. Consensus PCR with previously described primers (4) was used to amplify a 5-kb product from killer

whale genomic DNA (animal 0162). The reaction mixture included 200 nM each of primers P-tRNA and Pol-Cm (Table 1), 200 μM each of deoxynucleoside triphosphates (dNTPs), 1.8 mM of MgSO_4 (2 μl of buffer A and 8 μl of buffer B), 2 μl of Invitrogen elongase, and 200 ng of genomic DNA in a 50- μl reaction mixture. PCR was performed on an ABI 200 thermocycler as follows: (i) 30 s at 94°C ; (ii) 35 cycles, with 1 cycle consisting of 30 s at 94°C , 30 s at 65°C for 30 s, and 4 min at 68°C . PCR products were resolved on a 1% modified Tris-acetate-EDTA (TAE) agarose gel stained with crystal violet, cut out, and purified (Millipore ultrafree kit). PCR products were cloned using Invitrogen Topo XL. Sequencing was performed with the Beckman CEQ automated capillary sequencer and analyzed with the CEQ 8000 genetic analysis software.

Gene walking to obtain the 5' LTR. The SeeGene gene walking kit was used to obtain the region 5' of the 5-kb amplicon from the original consensus PCR (Fig. 1). In the first round of PCR, the template-specific primer KWERV 3 Walk1 (Table 1) was used with 100 ng of genomic DNA from animal 0162 with the kit's ACP primers. The second PCR included 3 μl of product from the first PCR, 250 nM KWERV 3 Walk2 (Table 1) and 2 μl of DW2-ACPN supplied in the kit. The third reaction consisted of 1 μl of product from the second PCR, 250 nM KWERV 3 Walk3 (Table 1) and 1 μl of the UniP2 supplied in the kit. PCR products were resolved on a 1.5% modified TAE agarose gel stained with ethidium bromide. PCR products were purified, cloned, and sequenced.

Amplification of env. A forward primer (3' Pol) was designed to the 3' end of the polymerase region and used with LTR 3' End, a specific primer designed to the U5 region of the LTR (Table 1), in order to amplify the envelope gene (Fig. 1). The reaction used Invitrogen elongase with the same conditions described for the original consensus PCR, but with a 56°C annealing temperature. The PCR product (40 μl) was run on a 1% modified TAE agarose gel stained with crystal violet, and an approximately 3.2-kb band was cut out and cloned using the Invitrogen Topo XL cloning kit. Sequencing confirmed the presence of the envelope gene.

Nearly full-length proviral amplification and cloning. Attempts to amplify the full-length provirus with both LTRs were unsuccessful due to preferential LTR amplification. A nearly full-length provirus without the 5' LTR was amplified initially from 200 ng of genomic DNA of brain tissue from animal 0162 (Fig. 2). Reaction conditions included 200 nM each of primers P-tRNA KW and KWERV 3' LTR (Table 1), 1.5 U Invitrogen Accuprime *Taq* high-fidelity DNA polymerase, 1 $\times$ Accuprime PCR buffer II, and 3 mM MgSO_4 (50- μl reaction mixture). PCR was run on an ABI 200 thermocycler as follows: (i) 94°C for 2 min; (ii) 35 cycles, with 1 cycle consisting of 20 s at 94°C , 30 s at 53°C , and 10 min at 68°C . Overlapping clones of the 5' end (1.5 kb) were produced with primers 5' LTR and KWERV *gag* R (R for reverse) using the same conditions, but with a 2-min extension step. Products were resolved on a 1% TAE agarose gel stained with ethidium bromide (Fig. 2), and the remaining 40 μl was resolved on a 1% TAE agarose gel stained with crystal violet. The nearly full-length product from the 200-ng reaction mixture and the 5' amplicon were cut out from the crystal violet gel and purified using the Invitrogen Topo XL purification system. Products were cloned using the Invitrogen Topo XL cloning kit, and six clones were sequenced in their entirety using the ABI 3100 prism automated sequencer. Nearly full-length products were also amplified from all killer whale genomic DNA samples extracted using the same reaction conditions and 200 ng of genomic DNA.

Sequence analysis. The cloned 5' end of KWERV and the nearly full-length amplicon were overlapped and assembled using Assemblylign software to generate a full-length proviral sequence. The overlapping region of the two cloned pieces was 1,044 bp in length.

Original searches were performed on the NCBI BLASTX database (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Initially, MacVector version 10 (37) was used for DNA and protein sequence alignments. Open reading frames (ORFs) were predicted using the NCBI ORF finder (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) in conjunction with comparative alignments to known retroviral ORFs. Phylogenetic analysis was performed comparing KWERV with several full-length gammaretroviruses: Moloney murine leukemia virus (Moloney MLV) (GenBank accession no. J02255 [45]), Friend MLV (GenBank accession no. M93134 [30]), xenotropic MLV-related virus (GenBank accession no. DQ399707 [55]), FeLV (GenBank accession no. M18247 [13]), porcine endogenous retrovirus A (PERV A) (GenBank accession no. AJ293656 [2]), PERV B (GenBank accession no. AY099324 [20]), PERV C (GenBank accession no. EF133960 [24]), *Mus dunni* endogenous virus (MDEV) (GenBank accession no. AF053745 [59]), koala retrovirus (KoRV) (GenBank accession no. AF151794 [15]), gibbon ape leukemia virus (GALV) (GenBank accession no. NC_001885 [9]), and RD114 (GenBank accession no. EU030001 [K. Ghani, M.-C. Caron, and M. Caruso, unpublished data]). Pairwise distances (P-distance model) were computed for nucleotide and amino acid alignments using MEGA3.1 software (21),

TABLE 1. Primers used in molecular characterization of killer whale endogenous retrovirus

Primer ^a	Sequence ^b	Use(s) ^c
P-tRNA	5'-CAD KTG GGG GCT CGT CCG GGA T-3'	Initial consensus PCR (4)
Pol-Cm	5'-GGG TYT CRA GTC CAG TCC ATC TTT CTT ACT T-3'	Initial consensus PCR (4)
KWERV 3' Walk1	5'-GAA ACC GAG TGA AGG GGA GAC AG-3'	Gene walking
KWERV 3' Walk2	5'-GTG AAG GGG AGA CAG AGT TTG CC-3'	Gene walking
KWERV 3' Walk3	5'-GCC AGA GGT CCG ATA CAG ATT CC-3'	Gene walking
3' <i>pol</i>	5'-AGG AAC TTA CTG GGA GGT GGA C-3'	<i>env</i> amplification
LTR 3' End	5'-CGC CCA GGA AAC AAC AAG AG-3'	<i>env</i> amplification
P-tRNA KW	5'-CAT TTG GGG GCT CGT CCG G-3'	nFL provirus
KWERV 3' LTR	5'-TGT TAG ACC TTC GCG GTC TCC TAG G-3'	nFL provirus
5' LTR	5'-TGT GGG ACT CGA GGG ACA TTG	5' provirus
KWERV <i>gag</i> F	5'-GGA CAT CCA GAC CAA GTA CCA TA-3'	<i>gag</i> RT-PCR, PCR
KWERV <i>gag</i> R	5'-GGC CAT ATT CTC TAA GAG GTA G-3'	<i>gag</i> RT-PCR, PCR
<i>pol</i> 2 F	5'-ATC ATT CAC TGC CTT GGA CAC C-3'	<i>pol</i> RT-PCR
<i>pol</i> 2 R	5'-TGA CTT CTC CCG ATG CCA TTT C-3'	<i>pol</i> RT-PCR
<i>env</i> 2 F	5'-GGA GAT ACT TAT TGG AGA CCC ACC TC-3'	<i>env</i> RT-PCR
<i>env</i> 2 R	5'-ACA TTC AGG GGC GAC GGG-3'	<i>env</i> RT-PCR
<i>env</i> 3 F	5'-ACC GGG GAT ACT AAT TGG AAA CC-3'	<i>env</i> RT-PCR
<i>env</i> 3 R	5'-TCC TGA ACA TTT ACT CCG GTG TCC-3'	<i>env</i> RT-PCR
KWERV <i>pol</i> F	5'-TGG GCA GCT GCG CTA CCA-3'	<i>pol</i> PCR
KWERV <i>pol</i> R	5'-CCA CCT CCC AGT AAG TTC CTG-3'	<i>pol</i> PCR
KWERV <i>env</i> 1Q F	5'-CCT CAC GTT GGC CAT GCT-3'	<i>env</i> PCR
KWERV <i>env</i> R1	5'-TCC CTC GAG TCC CAC ATT-3'	<i>env</i> PCR
<i>gag</i> 1 Q F	5'-AAC CCC AAC ACG GAG AAG CT-3'	<i>gag</i> qPCR
<i>gag</i> 1 P	5'-FAM- CGC CCC GGA TCC ACC GAA G -TAMRA-3'	<i>gag</i> qPCR probe
<i>gag</i> 1 Q R	5'-TGC GGG TCA GGA TAA ATC G-3'	<i>gag</i> qPCR
<i>gag</i> 2 Q F	5'-AAC CCC AAC ACG GAA AAG C-3'	<i>gag</i> qPCR
<i>gag</i> 2 P	5'-FAM-ACC CGG CGC CCC AGA TCC-TAMRA-3'	<i>gag</i> qPCR probe
<i>gag</i> 2 Q R	5'-TGC GGG TCA GGA TAA ATC G-3'	<i>gag</i> qPCR
<i>pol</i> 2 Q F	5'-CAT CTC TAC GGG TTA ATG TCA CAG A-3'	<i>pol</i> qPCR
<i>pol</i> 2 P	5'-FAM-TGG GTA ACT CGG TTC CCG CAA GC-TAMRA-3'	<i>pol</i> qPCR probe
<i>pol</i> 2 Q R	5'-CCC CAT ACC AGC GGT TTC T-3'	<i>pol</i> qPCR
<i>env</i> 2 Q F	5'-GAG GGA AAA ACA AGA AGG AAG AGA-3'	<i>env</i> qPCR
<i>env</i> 2 P	5'-FAM-CCT GTA ATC CTC ACA TTG GCT GTC-TAMRA-3'	<i>env</i> qPCR
<i>env</i> 2 Q R	5'-CCG GCA GAT ATT CTT AGT CCT AGA A-3'	<i>env</i> qPCR
IL-10 Q F	5'-TGC AAA ACC AAA CCA CAA GTC T-3'	qPCR single-copy std
IL-10 P	5'-VIC-ACT GAA CGA AGA AGG C-NFQ-3' (MGB labeled)	qPCR single-copy std
IL-10 Q R	5'-GCT GGG CAT GGT GGT GTT-3'	qPCR single-copy std
<i>gag</i> F SS1	5'-CCG AAA GCG ATT TAT CCT GAC-3'	Single-cell PCR
<i>gag</i> R SS1	5'-TGA CGA GTG GGA AAC CAA CAT-3'	Single-cell PCR
<i>gag</i> F SS2	5'-TGC TTC TTT TGG ATT CCC CAC-3'	Single-cell PCR
<i>gag</i> R SS2	5'-TGG AGA GGT CTA TTC CCT TGG-3'	Single-cell PCR

^a The letter F in a primer name (usually but not always at the end) indicates that it is a forward primer, while the letter R in a primer name indicates that it is a reverse primer.

^b Abbreviations: FAM, 6-carboxyfluorescein; TAMRA, 6-carboxytetramethylrhodamine; MGB, minor groove binder.

^c Abbreviations: nFL provirus, nearly full-length provirus; qPCR, quantitative PCR; std, standard.

FIG. 1. Strategy for obtaining sequence of the full KWERV proviral genome. Initial amplification was with consensus P-tRNA and Pol-Cm primers. Gene walking was achieved using three primers from the 5' region of the *gag* gene to acquire the 5' LTR sequence. A reverse primer designed for the U5 region of the genome (LTR 3' End) was used with a forward *pol* primer (3' *pol*) to amplify the remainder of the genome.

and neighbor-joining phylogenies were created for *gag*, *pol*, *env*, and the full-length genome. Because the envelope sequence of the only natural reported full-length GALV is truncated, we compared our results for the *env* gene with those obtained using the GALV SEATO envelope (GenBank accession no. AF055060 [52]). For analysis of the full-length genome based on protein sequence, putative amino acid sequences of Gag, Pol, and Env were concatenated for each virus. Distances were tested with 1,000 bootstrap replications. Nucleotide alignments for *gag*, *pol*, *env*, and the full-length genome were performed using ClustalW 1.83 (7), and Bayesian phylogenies were computed using MrBayes v3.1 (40). For each gene and also the full-length genome, the Markov chain Monte Carlo was run for 100,000 generations, sampling every 100th generation after a burn-in of 250 generations. Neighbor-joining trees of nucleotide sequences and amino acid sequences were compared with Bayesian trees to determine the most probable results (see Fig. 4). Splice donor and acceptor sites were predicted using the splice predictor tool available through the Berkeley Drosophila Genome Project (<http://www.fruitfly.org>).

Reverse transcription-PCR (RT-PCR). Total RNA from tissue and blood samples was used to amplify regions of *gag*, *pol*, and *env*. Reaction conditions for the *pol* and *env* genes included 200 μ M dNTP, Amplitaq gold buffer, 250 μ M MgCl₂, 1.25 U Amplitaq gold, and 500 nM each of primer pairs *pol* 2 F/*pol* 2 R (F for forward primer and R for reverse primer), *env* 2 F/*env* 2 R, and *env* 3 F/*env* 3 R (Table 1). Reactions were performed with 1 μ l of cDNA on the ABI 200

FIG. 2. Optimized PCR at 35 cycles for amplifying the nearly full-length KWERV from genomic DNA of a killer whale (animal 0162). The expected product size for the nearly full-length genome without the 5' LTR is approximately 7.8 kb. The four lanes contain increasing concentrations of starting total DNA (50 to 200 ng). The leftmost lane contains molecular size markers. NTC (-), no-template control (negative control).

thermocycler as follows: (i) 5 min at 94°C; (ii) 35 cycles, with 1 cycle consisting of 1 min at 94°C, 1 min at 55°C, and 30 s at 72°C; and (iii) a final extension step of 7 min at 72°C. For KWERV *gag*, the PCR conditions were the same but with an annealing temperature of 55°C. For *gag* amplification, a subsequent reaction was performed using 1 μ l of PCR product with the same conditions. Controls with no added reverse transcriptase were run in parallel. Primer sequences are listed in Table 1. All PCR products were purified with the Millipore ultrafree kit and cloned using Invitrogen Topo TA cloning kit. Cloned products were sequenced to verify that their origin was KWERV.

Copy number analysis. To determine KWERV gene copy numbers in the killer whale genome, quantitative PCR-based TaqMan assays were developed for *gag*, *pol*, and *env* using an *Orcinus orca* interleukin-10 (IL-10) sequence (GenBank accession no. U93260) as a single-copy standard in duplex reactions. Primer/probe sets were designed using the ABI Primer Express software (version 3.0; Applied Biosystems, Foster City, CA) (Table 1). Each reaction mixture consisted of 10 ng of genomic DNA, 900 nM each of the retroviral forward and reverse primers, 150 nM each of the IL-10 forward and reverse primers, and 250 nM each of the IL-10 and retroviral probes. Reactions were in a volume of 20 μ l, with 10 μ l consisting of Applied Biosystems TaqMan Universal master mix with AmpErase. Four replicates of each sample were run, along with four replicates of the no-template controls. The TaqMan program included 1 cycle of 2 min at 50°C, 1 cycle of 10 min at 95°C, and then 40 cycles, with 1 cycle consisting of 15 s at 95°C and 1 min at 60°C. Results were analyzed with the threshold cycle (C_T) set at 0.2 using the ABI TaqMan copy number macro and ABI CopyCaller software (version 1.0), which computed estimated copy numbers.

Southern hybridization. Genomic DNA (10 μ g) from brain and spleen tissue of two unrelated captive killer whales and also liver tissue from a beluga whale was digested with 10 U of EcoRI and PstI at 37°C overnight. DNA was run on a 1% TAE agarose gel and stained with ethidium bromide to visualize DNA digestion. DNA was denatured by soaking the gel in 1.5 M NaCl with 0.5 N NaOH and neutralized in 1 M Tris (pH 7.4) with 1.5 M NaCl, transferred to a nylon membrane by capillary transfer in 10 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate) for 17 h and fixed to the membrane using UV irradiation. A 32 P-labeled probe was prepared with the Roche nick translation kit using a 410-bp *gag* PCR fragment amplified from a KWERV plasmid using the primers KWERV *gag* F and KWERV *gag* R (Table 1). Southern hybridization was performed using previously described conditions and methods (42) and a phosphorimager.

Analysis for KWERV in other species. All samples used for this work were collected opportunistically from captive animals. Livers were collected from one each of a common dolphin (*Delphinus delphis*), a Commerson dolphin (*Cephalorhynchus commersonii*), two bottlenose dolphins (*Tursiops truncatus*), a false killer whale (*Pseudorca crassidens*), a Risso's dolphin (*Grampus griseus*), a rough-toothed dolphin (*Steno bredanensis*), a Pacific white-sided dolphin (*Lagenorhynchus obliquidens*), two beluga whales (*Delphinapterus leucas*), a fin whale (*Balaenoptera physalus*), three harbor porpoises (*Phocoena phocoena*), a dwarf sperm whale (*Kogia sima*), a pygmy sperm whale (*Kogia breviceps*), and a hippopotamus (*Hippopotamus amphibius*). Blood samples were collected from two short-finned pilot whales (*Globicephala macrorhynchus*) into tubes containing EDTA and frozen at -80°C until extraction. Kidney tissues from a domestic pig (*Sus scrofa*) and a mouse (*Mus musculus*) were used as controls. Genomic DNA was extracted as described previously using the Qiagen DNeasy kit. PCR was performed with primer pairs KWERV *gag* F/KWERV *gag* R, KWERV *pol* F/KWERV *pol* R, and KWERV *env* 1 QF/KWERV *env* 1 R1 (Table 1). Reaction conditions for each gene included 200 μ M dNTP, Amplitaq buffer with 150 μ M MgCl₂, 1.25 U of Amplitaq, and 1 μ M each of primers. Reactions were performed with 100 to 200 ng of genomic DNA as follows: (i) 1 cycle of 10 min at 94°C; (ii) 35 cycles, with 1 cycle consisting of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C; and (iii) a final extension step of 7 min at 72°C. PCR products were resolved on a 1.5% TAE agarose gel and purified using a Qiagen MiniElute column. Products were sequenced in both directions using the ABI 3100 prism capillary sequencer to confirm positive results. In addition, an Ensembl database search (<http://www.ensembl.org/index.html>) for full-length KWERV in the recently sequenced bottlenose dolphin (*Tursiops truncatus*) genome (<http://www.hgsc.bcm.tmc.edu/project-species-m-Dolphin.hgsc?pageLocation=Dolphin>) was performed, and results were compared to the original KWERV sequence.

Sperm single-cell PCR. Sperm collected from a live captive killer whale was diluted in phosphate-buffered saline (pH 7.4) to one cell per 0.5 μ l. One cell (0.5 μ l) was transferred to each of 21 PCR tubes and visualized under a microscope to confirm transfer. Viagen Biotech DirectPCR lysis buffer (1.5 μ l) with 0.1 mg/ml proteinase K was added to each tube, and the sperm were incubated at 55°C for 3 h and then lysed at 85°C for 45 min. Conventional PCR was performed on four of the samples using two rounds of nested PCR. The first round included 200 μ M dNTP, Amplitaq gold buffer, 250 μ M MgCl₂, 1.25 U Amplitaq gold, and 100 nM each of primer pairs *gag* F SS1 and *gag* R SS1. Reactions were run as follows: (i) 5 min at 94°C; (ii) 45 cycles, with 1 cycle consisting of 30 s at 94°C, 30 s at 58°C, and 45 s at 72°C; (iii) a 7-min final extension step at 72°C. Nested PCR was performed using primers *gag* F SS2 and *gag* R SS2 under similar conditions, except that the primer concentrations per reaction were 500 nM each and the extension time was 45 s per cycle.

As an adjunct to conventional PCR, TaqMan PCR was performed as a sensitive means to detect KWERV *pol* DNA on a nonquantitative basis. TaqMan conditions were the same as described for the copy number assays, except that no IL-10 primers or probe were included. Positive controls included replicates of 10 pg of killer whale genomic DNA.

Nucleotide sequence accession number. The most intact assembled provirus sequence reported here is recorded in the GenBank nucleotide sequence database under accession no. GQ222416.

RESULTS

Molecular characterization. Initially, 5 kb of KWERV sequence was amplified from DNA from multiple tissues of the index animal (animal 0162) with retroviral consensus PCR primers (4) (Fig. 1). The 436-bp 5' LTR was obtained through gene walking. The remainder of the genome was amplified and sequenced based on primers designed against the 3' end of the LTR.

Sequences of KWERV were detected in DNA from peripheral blood mononuclear cells and tissue samples of all killer whale samples examined using a variety of specific primers listed in Table 1. PCR conditions were optimized for nearly full-length proviral sequences (not including the 5' LTR) (Fig. 2) using KWERV primers P-tRNA KW (coding) and KWERV 3'LTR (noncoding). Nearly full-length sequences were documented in all the animals. Six nearly full-length clones from the same animal (0162) were sequenced. Four out of the six clones

FIG. 3. Full-length proviral sequence for killer whale endogenous retrovirus sequenced from killer whale (KWERV). The predicted divisions of LTRs (U3, R, and U5) are denoted by arrowheads. The features of the sequence are contained in boxes, including the CCAAT box, the TATA box, the polyadenylation site (PAS), the primer-binding site (PBS), the Cys-His box, the polypurine tract (PPT), the predicted splice donor site (SD), and the predicted splice acceptor site (SA). The boundaries of the three genes, *gag*, *pol*, and *env*, are denoted by arrows. The stop and start codons found in the polymerase gene between the RNase H and integrase are shown in boldface type. Note that the nearly full-length sequence beginning with the primer-binding site was from one clone, and the reported 5' LTR was from an overlapping clone of the 5' end of the genome extending approximately 1 kb into the leader and *gag* sequences.

```

4501 GGGCACTGGA GATGGCAGAG AATAAAAAGG TAAACATCTA CACGGACAGT AGATATGCAT TTGCTACCGC TCATATCCAC GGTACCATCT ACCAACAGAG
4601 AGGGTTGCTA ACCTCAAGTG GCAGAGAAAT TAAAAACAAA GACGAAATTTG TGGCTCTATT GGTTCGACTT ATGCTTCCCA CTAAAGTTAG CATCATTCAC
4701 TGCCCTGGAC ACCAGAAAGA CAATACCCCA ATAACATAGG GTAATAATAT GGCTGACAAG GTTGCCTGAG AAACGGCATC GGGAGAAGTC ATTTTAGGAC
4801 TGTCAGATAA AGTTCCTGAA AAACCAGGCA GCGACGAGGA AATCATCCCG ACCACAACCA AAGGATCTTT GTCCCCTCAG CAAGCAGAAT CCATGTTACC
4901 ACAGATTTCAT AGATGGACAC ACCTGGGGAC TAAAAAATG ATATCCCTGT TACATAAAAC CGGGTACGGA ACCCCTGGAT TGACAAAATT AGCTGAACAA
5001 ATTGTACAGG AATGTGTTCC ATGTCAACAG GTAATGCTC ACAGGGGAAA ACTTGAAACT GGAAAAAGGC TCAGGGGAGA CCACCCAGGA ACTTACTGGG
5101 AGGTGGACTT TACCGAAGTA CGTCTGGAA GGTACGGTAA TAAATACCTC CTGGTTTTTA TAGATACTTT TTCAGGATGG ATAGAAGCCT TCCCAACGAA
5201 GAAAGAAACA GCTGCTGTGG TAGCCAAAAA GATCCTAGAA GAAATTTTTC CTCGATTGTT GGCACCAAGG GTAATAGGGT CCGACAATGG TCCGGCCTTC
5301 GTCGCCCAGG TAAGTCAGGA TGTGGCCAGA TATTGGGGA CTGATTGGAA ATTACATTGT GCATATAGAC CCCAGAGTTC AGGTCAAGTA GAAAGAATGA
5401 ATAGAACTCT AAAAGAGACC CTAACATAAT TGTCTTGGG GACTGGCAGT ACAGATTGGA CGGTGCTCCT TCCTTTAGCC CTGTTCCGGG TTAGGAATAC

5501 ACCTTCCCGG TACCATCTTA CTCCTTTTGA GGTGCTATAC GGGGCCCCAC CTCCCCTTCT CACCTTAGGA GAAGAAATAA AACCAGACTG TCAAAATAAC
                                     ↓SA
                                     →Env
5601 ACTGACTTGT ATGCTAGGCT ATTGGGACTC CAACTGGTCC AAAAGGAGAT ATGGTCCCAA CTCGCCGAGG CCTACCAACC GGGAACACCA GGAGAACTC
5701 ATCTTTTCCA AGTCGGAGAC TCCGTATACG TCCGTCGGCA TCGAACCAG ACGTGGAAC CTCGATGGAA AGGACCATAC ACCGTCCTCC TCACCACCCC
5801 AACGGCCATA AAAGTGAGC GCATCGCTGC CTGGATTTCAT GCTTCACAGC TGAAAGCTGC ACCAGCGACG GCATCATCAG GATGGTGGG CCAAAGAACC

                                     Pol←
5901 GACAACCTGC TCAAACCTCA CGTTCTACGA ACCTAAGACT TATAATTTTG ATTTTACTGC CTTTATTGAC TGTAAGTGAT TTTAGCCCTC ACCTGCCCCA
6001 ACGTCTAACT TGGCAGGTAA TTTCTCAAAC TGGAGGTGTA ACTTGGTCCA TTAGCCATAC TGCACCTCCC TGGACCTGGT GGCCAGACCT TTTTCTGAT
6101 GTCTGTAAAC TGGCCATAGG AGCTCCCTAT TGGGATCTAG AGGGTTATTA TGATCAGCAT AATGCTCCGT CCGAACTTCT AACTAGAGCT GATTACTCTG
6201 GAGAGGGTTG TAAAAATCAG GTCGGAAGAA GTTGCTCAG AACCTCTCC TTCTATGTAT GCCCGGGTT CCATCGCCCC CGATCCCTGA ATTATAATG
6301 TGGGGGAAGT GAAAATTTTT ATTGCCAAAA CTGGGGATGT GAAACCACAG GAGATACTTA TTGGAGACCC ACCTCAACTT GGGATTTTAT TACAGTAACA
6401 GCCAACTATA CTCATACTTC TTATAGGGGT CCCCACCCG TCGCCCTGTA ATGTTCAAGG TGGTGCCAGC CCTTCAAAT TTCCTTTAGT AATCCTGGAA
6501 AAAGGGATAA AGAATGGATA AATGGCCATT CATGGGGCAT TAGATTTTAC AAGAATGGAT ATGATTCGGG AATATTGATG ACCCTTAAGC TAAAAATTGA
6601 GACTCCCCTC CTGTGTGCA TAGGCCCAA CCCTGTACTT GGCCCCCCC CGTTACCCCC AAACCTGTA CTTGGCGCCC CCCTGTTACC CCCAGCCCCT
6701 AAAAATAACG GGACCTCTGG GCTGAAAGAG ACAACGATTA AGCCCGGAAC CCGCCAGGAT AGAATGCTTT CTTTGGTGCA GGCAGCTTTT ACGGTCCTTA
6801 ATGCTACAAA CCCGGAGGCT ACAAATCAT GTTGCTCTG TTATGCTGCC CCTCCCCCTT ATTATGATGC TATAGGATAT AGTTCTAATT ATACTAATGT
6901 CAGCTCCCCG GACCAATTGC GATGGAAACA GGAAGGGAAC AGTAAATTAA CTCTGTCTTC GGTTTTGGGA AATGGTACTT GTGTAGGAAC ACCTCCCCAG
7001 TCCCATCGAC ACCTTTGTCA TGATTTTAGC CTCCTTCAG GCTCAGGATA TCTTGTTCCT CCCCAGATG GATGGTGGG ATGTAACACG GGGCTAACCC
7101 CTTGCGTTTC TCTAGAAGTT CTCAACACTT CAGCTGATT TGTGTACTA GTGCAATTAG TTCCTAGACT TATCTATCAT TCAGATCCGT CCTTCTGGGA
7201 TGAATACGAG GGAAGAAAAA AGAAAAAGAT AACCTCACG TTGGCCATGC TCCTAGGACT AGGAATAACT GCTGGGATAG GAACAGGAAC CACAGCCCTC
7301 ATACAACAAC CCCAGTATTA TGCAAGTTTA AGACAGGCTG TAGACATCGA CCTTCGAGCA TTAGAAAGTT CCATAACTCA ACTAAAGGAA TCACCTACCT
7401 CACTTTCAGA AGTGGTGTG CAGAATAGAA GAGGCCTAGA TTTGCTGTTT CTTAAGGAAG GGGGATTATG CGCCGCTCTA AAAGAAGAAT GTTGCTTTTA
7501 TATTGATCAT TCAGGAACCA TTACCAAAAC CATGGATAAA CTCAGGGAAC GCTTAGATAA AAGGCAAAGG GAGAGAGAAA ACGAAAAAGG ATGGTTTCAA
7601 TCATGGTTTG ATAAATCCCC CTGGTTTACC ACCTTAATCT CTACTCTGTT GGGACCTCTT ATTGTTTTAT TGTGATTTT AACTTTTGA CCATGTATTC

7701 TAAATCGCTT GATAGCATTT ATTAGAGGAC GTATCAGTAC TGTACAAGTT CTAATGTCAA GACAACAGTA CCAAAGTTTA CGAACAGAAG ATTGAAATTC
                                     Env←

                                     PPT →3' LTR
7801 CATGATTGGA ATCAATAGTC ACAAGAAAAA GGGGAAATG TGGGACTCGA GGGACATTGT TCCAGCTAAT GATTAAGAAC TCTCCAGACA ATAGGGGCTT
7901 CTTAGACAAAT GGGTGAATTT CCGGGATGTC CGGCCCCCTT CCGAGAAGGA GGGAGATAAA CAAATGTAA AAAAGTATC TGTCAAAGAA CCAATCAGGA
8001 AGCCTGGGAC AGGTTCCTC TCTGGTCCGA ACAGAAGCTA AGAACCTATC ATTAATAATC ACAATTGAAT CCACACTTTC CATAACAGGA AGTTGAGACC
8101 TATAAAAATG CATGTTACG CCCTGGGGGG GTTCCTCTCT CGACCTPCGT CGTGAGGACC AAGGAACCCC GGTGCACCGG CCTTCAATAA ACCTCTGCG
8201 TTTTGATCG GCTTCCGACT CTTGTTGTTT CTGGGCGAG TCGAACTCC TAGGAGACCG CGAAGGTCTA ACA

```

FIG. 3—Continued.

had stop codons and frameshifts disrupting the open reading frames, while two identical clones (i.e., the sequence reported here) had complete ORFs with the exception of a single stop codon in the polymerase gene, discussed below. A predicted splice donor site is located at positions 502 to 503, and a predicted splice acceptor is found at bases 5567 to 5568.

The assembled full-length sequence of the most genomically intact KWERV has been deposited into GenBank (accession no. GQ222416), and our subsequent molecular analysis was focused upon this sequence. It is comprised of 8273 nucleotides (Fig. 3). GenBank indicated that the closest alignment was with porcine endogenous retrovirus, and phylogenetic analysis revealed a 57% identity at the nucleotide level and 61% identity at the amino acid level for all three ORFs (i.e.,

Gag, Pol, and Env) considered together. A comparison of the Bayesian phylogenetic tree of the full-length nucleotide sequences with the neighbor-joining tree of the amino acid sequences shows that KWERV clusters with MDEV, PERV, KoRV, and GALV (Fig. 4) though it is clearly in a distinct line. These results were also consistent for each gene individually using the same analyses.

From the six proviruses sequenced, there were two separate sets of LTRs identified for this endogenous retrovirus family. One set is identical on each end of the viral genome and is comprised of 436 bp, each of which is bounded by a 3-bp inverted repeat. The second LTR has a 2% divergence in sequence between its 3' and 5' ends, suggesting it has been integrated into the killer whale genome longer than the first

FIG. 4. Phylogenetic trees of full-length KWERV with other gammaretroviruses. (A and B) Bayesian tree was constructed based upon full-length retroviral nucleotide sequences with 100,000 generations (A) and compared to a neighbor-joining tree based upon amino acid sequences designed from concatenated Gag, Pol, and Env with 1,000 bootstraps (B). Clustering for KWERV appeared similar for both, with KWERV grouping among PERV, MDEV, GALV, and KoRV. PERV A (porcine endogenous retrovirus A) (GenBank accession no. AJ293656), PERV B (GenBank accession no. AY099324), PERV C (GenBank accession no. EF133960), MDEV (*Mus dunni* endogenous retrovirus) (GenBank accession no. AF053745), KoRV (koala retrovirus) (GenBank accession no. AF151794), GALV (gibbon ape leukemia virus) (GenBank accession no. NC_001885), KWERV (killer whale endogenous retrovirus) (GenBank accession no. GQ222416), FeLV (feline leukemia virus) (GenBank accession no. M18247), XMRV (xenotropic murine leukemia-related virus) (GenBank accession no. DQ399707), MMLV (Moloney murine leukemia virus) (GenBank accession no. J02255), FMLV (Friend murine leukemia virus) (GenBank accession no. M93134), RD114 (RD114 feline endogenous retrovirus) (GenBank accession no. EU030001). The bars for both trees show the number of substitutions per site.

LTR. It is approximately 89% identical with the first LTR. The 3' LTR is preceded by a 16-bp polypurine tract. The TATA box is located at position 263. The transcriptional start site is yet to be determined, but it is thought to fall around position 294 based upon alignment with the koala gammaretrovirus LTR (15). The area surrounding the putative KoRV transcriptional start site is 63% identical over 20 bp to KWERV. Therefore, the U3 region is estimated to be 293 bp. The R region is 76 bp spanning from the estimated transcriptional start site, past the poly(A) signal at position 348 to a CA dinucleotide located 16 bp downstream of the polyadenylation site. The U5 region spans positions 370 to 436, making it 67 bp in length. The CCAAT box (a common promoter motif in eukaryotes) is located at position 163. The 3' end of U5 in the 5' LTR is defined by the presence of an 18-bp proline tRNA primer-binding site, which primes plus-strand synthesis during reverse transcription.

The predicted start codon for *gag* (1,584 bp) is at position 849. The *gag* sequence aligns most closely with PERV: 62.3% identity at the nucleotide level and 62.1% at the amino acid level. A viable Cys-His box motif, commonly found in the nucleocapsid region and important in viral encapsidation, is located at position 2325. Notably, both the PPPY and PSAP L domains are intact. In contrast, the endogenous form of KoRV has a dysfunctional form of the PPTY motif, resulting in significantly lower levels of viral particle assembly. Thus, the reported titers for KoRV were approximately 1,800-fold lower than those for GALV (10, 34).

The predicted start codon for *pol* (3504 bp) is at position 2433, immediately after the leaky stop codon common to all gammaretroviruses (8). A stop codon is located between the RNase H and integrase domains at position 4767, but another start is found at the beginning of the integrase at position 4893. The significance of this feature is unknown, as gammaretroviruses do not typically present with a stop codon in this location.

However, they are known to read through stop codons in other locations, such as the one commonly found between the *gag* and *pol* genes (8). The polymerase region also aligns most closely with PERV (64% nucleotide identity and 67% amino acid identity).

The predicted start codon for *env* (2145 bp) is at position 5651. The *env* nucleotide sequence aligns most closely with the koala gammaretrovirus (54.2%). However, the amino acid sequence aligns most closely with a murine endogenous retrovirus (52.4%). A second KWERV envelope sequence (*env* B) was identified by amplification of the envelope gene alone, but it was not found with any of the full-length sequences. This variant is disrupted by stop codons at positions 6131 and 6509 and is 89.7% identical to the first at the nucleotide level and 86.6% identical at the amino acid level. It also aligns most closely with KoRV at the nucleotide level (54.5%) and with MDEV at the amino acid level (52.2%). A possible limitation to this first analysis was the use of a GALV *env* sequence that is truncated by several amino acids at the 3' end. Repetition of the analysis with the full-length GALV SEATO *env* sequence produced the same results (GenBank accession no. AF055060).

RNA expression. RT-PCR of multiple tissues (brain, axillary and cervical lymph nodes, adrenal gland, kidney, serum, skin, and spleen) from the original animal (animal 0162) showed *gag* expression in brain, axillary lymph node, and spleen. *pol* RNA expression was demonstrated in brain, axillary lymph node, spleen, and serum. Expression of the first *env* variant (*env* A) was demonstrated in cervical and axillary lymph nodes, spleen, adrenal gland, serum, and brain, while transcription of the second variant (*env* B) was present in the same tissues as well as skin. Out of 11 blood samples tested, 7 were RT-PCR positive for *env* in peripheral blood mononuclear cells (four for *env* A and all seven for *env* B), and 9 were RT-PCR positive for

FIG. 5. Calculated copy number results from tissue samples from three different killer whales. 0162 represents the original animal screened, while CKW represents an animal from a different institution. WKW1 represents a free-ranging killer whale. In all cases, copy numbers for *gag*, *pol*, and *env* appear relatively stable between tissues and animals, in most cases varying between two and four copies per diploid genome. C. LN, cervical lymph nodes; LN, lymph nodes.

pol, but no viral RNA was ever amplified from plasma (data not shown).

Quantitative copy number assays and Southern hybridization. Quantitative TaqMan PCR-based assays for proviral copy numbers were devised for *gag*, *pol*, and *env* and tested in multiple tissues in animal 0162. Subsequently, tissue samples from two additional animals were tested using the same assay and compared (Fig. 5). In all cases, the copy numbers were between two and three and did not exceed four copies per diploid genome. Results for Southern blots are shown in Fig. 6. Restriction enzyme PstI was chosen for digestion due to its placement with respect to the *gag* probe, and EcoRI (which has no sites in the reported sequence) was added for more complete digestion of flanking sequences. Genomic DNA from tissues of two captive animals from different facilities was compared. Banding patterns appeared the same for all samples, demonstrating a common integration site and supporting the conclusion that KWERV is an endogenous retrovirus present in all killer whales.

Analysis for KWERV in other species. Results for PCR analysis in relation to the phylogeny of other cetacean and artiodactylid (i.e., hoofed mammals) species are shown in Fig. 7. The domestic pig, hippopotamus, fin whale, and beluga whales were all negative for KWERV *gag*, *pol*, and *env*. All delphinoid species tested were PCR positive for all three KWERV genes, *gag*, *pol*, and *env*. In addition, both animals from the *Kogia* genus (dwarf and pygmy sperm whales) and *Phocoena* genus (harbor porpoises) were positive for *gag*, but not *pol* and *env*. Sequencing of the *gag* products from these species showed that the amplicons were closely related but distinct retroviral sequences from KWERV (70% identical for *Kogia* and 81% identical for *Phocoena*). In contrast, *gag* products from all the delphinids were nearly identical (95 to 98%) to the original KWERV nucleotide sequence. The Ensembl database search for full-length KWERV in the bottlenose dolphin (*Tursiops truncatus*) genome yielded two full-length sequences (scaffold 2349 and scaffold 111015) that were 97.4% and 96.9% identical to the KWERV nucleotide sequence, re-

FIG. 6. Southern hybridization of PstI- and EcoRI-digested genomic DNA using a ³²P-radiolabeled *gag* probe. The banding patterns for each digest were the same for lymphoid tissue and brain samples from two animals from different facilities, suggesting that sites of integration are the same between samples. 0162 represents the original animal screened, while CKW represents an animal from a different institution. A liver sample from a beluga whale was run as a negative control (-).

spectively. However, both *Tursiops* sequences were disrupted by numerous frameshifts and stop codons.

Single-cell PCR of sperm. Through conventional nested PCR, two of four single sperm were PCR positive for KWERV

gag. The TaqMan *pol* assay showed 10 of 17 single sperm as positive, with a median *C_T* of 36.3 (with a positive control of 10 pg of genomic DNA *C_T* value of 29.3 and no-template controls completely negative). As a frame of reference, human sperm have been reported to have an average of 3 pg of DNA per cell (43).

DISCUSSION

We conclude from the data reported here that the retrovirus initially amplified from killer whales exists in multiple odontocete species (Fig. 7) and that KWERV should be classified among the gammaretroviruses. The KWERV genome exhibits all the classical features of a gammaretrovirus, including a CCAAT box, a TATA box, a Cys-His box, a polyadenylation signal, a proline-based tRNA primer-binding site, splice donor and acceptor sites, the polypurine tract, and the PPXY/PSAP L domains (8). We also conclude that the virus is endogenous in killer whales and likely other delphinids based upon the ubiquitous presence of the provirus among individuals of the species, consistent copy numbers in individuals and tissues, common integration sites, and presence in single sperm cells of a killer whale. However, it is quite possible that exogenous variants of the virus exist in nature. On the basis of the results of phylogenetic analysis of KWERV and other gammaretroviruses, it groups most closely with the PERV/GALV clade (Fig. 4). There are both endogenous and exogenous members of this retrovirus family. While some have been directly linked to disease in their respective species (e.g., GALV), others are in

FIG. 7. Phylogenetic distribution of KWERV. The families and species in boldface type represent members tested for *gag*, *pol*, and *env* genes of KWERV using PCR. Those with a plus sign in parentheses were positive for at least one KWERV gene by PCR. Those with a minus sign in parentheses were negative for all KWERV genes. Sequencing of *gag* products from Kogiidae and Phocoenidae showed a related *gag* gene with 70% and 81% nucleotide identity to KWERV, respectively. Sequencing of *gag* products from Delphinidae showed nucleotide identities to KWERV ranging from 95 to 98%.

the process of being endogenized (e.g., KoRV), and many appear to be innocuous to the host (e.g., PERV).

The closest documented relative of KWERV is PERV. However, the percent identity at both the nucleotide and amino acid level is not substantially different than the average for the computed pairwise distances of all gammaretroviruses included in the analysis (57% nucleotide identity with an average of 59.2% and 61% amino acid identity versus an average of 60.6%, respectively). It is apparent that the most recent viral ancestors from which KWERV evolved are not presently known. Of note is the fact that pigs and cetaceans diverged from a common ancestor over 60 million years ago (1) (Fig. 7), but PERV is estimated to be only about 7 million years old (53). Thus, although pigs and cetaceans share a common ancestor, KWERV was not derived from an endogenous precursor in that common ancestor. It follows this conclusion that not all cetaceans are positive for the virus (e.g., fin and beluga whales). Thus, KWERV appears to have been endogenized after the divergence between cetaceans and artiodactylids (i.e., hoofed mammals).

On the basis of our initial characterization of KWERV in killer whales, this retrovirus is not a disease-causing agent for this species, as there is currently no molecular evidence for productive infection *in vivo*. Stable copy numbers of only two to four suggest that the virus either became replication defective quickly after integration into killer whales or that the species quickly adapted to prevent further viral replication. While the virus is transcriptionally active in multiple individual animals and in multiple tissues, viral RNA was not found in cell-free plasma from any individual animal. Viral polymerase and envelope RNA was amplified from serum of the original animal (animal 0162), but TaqMan copy number results from that animal were the same in all the tissues tested. It is possible that the positive result from the serum was due to cellular RNA contamination. Alternatively, endogenous retroviral RNA and proteins are produced routinely *in vivo* without the production of replication-competent viral particles (56, 57, 61). In addition, attempts to visualize the virus through electron microscopy or express the virus *in vitro* through activation of peripheral blood mononuclear cells have not been successful thus far (data not shown).

On the basis of the nearly intact nature of the open reading frames, the transcriptional activity of the provirus, and the identical LTRs flanking one of the two variants, it appears that at least one of the variants found in the killer whale was a relatively recent integration into the genome. Endogenous retroviruses can be useful markers for reinforcing evolutionary relationships (22, 32, 61). Delphinids are estimated to have diverged close to 12 million years ago (1, 5), and it is tempting to hypothesize that this retrovirus was endogenized before that time by a delphinoid ancestor, since all delphinoid species examined were positive for all three retroviral genes (Fig. 7). However, while analysis of the bottlenose dolphin genome showed a provirus with 97.4% nucleotide identity to this retrovirus, the ORFs were severely disrupted by stop codons and frameshifts, suggesting it was integrated into that species at an earlier point in time than the most intact KWERV sequence we obtained. In addition, positive PCR results were obtained for a related *gag* gene in both harbor porpoises and in *Kogia* species, but not in beluga whales. These results are inconsistent

with the phylogenetic groupings reported for cetaceans thus far (1, 5, 23, 33), suggesting that related exogenous retroviruses have infected several odontocete species and endogenized independently at different times.

It is unknown at this time whether an exogenous variant of this virus is still circulating in cetacean populations. It is noteworthy that harbor porpoises, a dwarf sperm whale, and a pygmy sperm whale were PCR positive for the *gag* gene, but not for *pol* or *env*, suggesting the presence of either a severely disrupted provirus or an altogether different but related virus. While the *gag* portion amplified from dwarf and pygmy sperm whales and harbor porpoises was related to KWERV, the sequences were far more divergent from KWERV than those amplified within the Delphinidae family. There are two possible explanations for the presence of the *gag* gene in these species. Either a KWERV predecessor was incorporated into the cetacean genomes before the divergence of those families (16 to 19 million years ago for harbor porpoises, 32 million years ago for *Kogia* [1, 5]) and has coevolved with its respective hosts, or a different (but related) retrovirus was integrated into the genomes of those families independently. The failure to amplify any KWERV-related sequences from beluga whales argues against the first hypothesis, based upon the estimated divergences of those species (Fig. 7).

Viral diseases in cetaceans have been largely unexplored, and it is imperative to gain understanding of the infectious diseases affecting these species in order to guide decisions for managing husbandry and veterinary medical care. It is also becoming increasingly important to further an understanding of the environmental and genetic factors affecting the onset and course of disease in free-ranging populations of marine mammals. Although the study of retroviruses is a unique challenge in these animals, their proclivity for jumping species unpredictably (26, 27, 54, 55) makes them an important area of focus. Moreover, the clear homology between this novel retrovirus of cetaceans and the gammaretroviruses of terrestrial mammals suggests that the histories and life cycles of these retroviruses have intersected and may continue to do so.

ACKNOWLEDGMENTS

We thank Tammy Tucker, Erika Nilson, and Frank Harrison for excellent technical assistance and Pauline Lee and Stephanie Cherqui for invaluable technical advice. We also thank Carolyn Wilson for advice and helpful comments on the manuscript. We thank James Casey, Davey Smith, Douglas Richman, Ellen Sparger, Marcy Auerbach, and Katie Marcucci for all of their intellectual input. We also thank Stephen Raverty, Ted Cranford, Megan Stolen, Wendy Noke, and Todd Robeck for their assistance with the collection of samples.

This work was funded by the following: NIH R01 AI52349 (D.R.S. and S.A.L.), unrestricted research support from the Busch Entertainment Corporation (S.A.L.), NIH training grant DK007022 (S.A.L.), the Molly Baber Research Fund (D.R.S. and S.A.L.), and the Verna Harrah Research Fund (D.R.S. and S.A.L.).

REFERENCES

1. Arnason, U., A. Gullberg, and A. Janke. 2004. Mitogenomic analyses provide new insights into cetacean origin and evolution. *Gene* 333:27–34.
2. Bartosch, B., R. A. Weiss, and Y. Takeuchi. 2002. PCR-based cloning and immunocytological titration of infectious porcine endogenous retrovirus subgroup A and B. *J. Gen. Virol.* 83:2231–2240.
3. Bossart, G. D., R. Ewing, A. J. Herron, C. Cray, B. Mase, S. J. Decker, J. W. Alexander, and N. H. Altman. 1997. Immunoblastic malignant lymphoma in dolphins: histologic, ultrastructural, and immunohistochemical features. *J. Vet. Diagn. Invest.* 9:454–458.

4. Burmeister, T., S. Schwartz, and E. Thiel. 2001. A PCR primer system for detecting oncoretroviruses based on conserved DNA sequence motifs of animal retroviruses and its application to human leukaemias and lymphomas. *J. Gen. Virol.* **82**:2205–2213.
5. Caballero, S., J. Jackson, A. A. Mignucci-Giannoni, H. Barrios-Garrido, S. Beltran-Pedrerros, M. A. Montiel-Villalobos, K. M. Robertson, and C. S. Baker. 2008. Molecular systematics of South American dolphins *Sotalia*: sister taxa determination and phylogenetic relationships, with insights into a multi-locus phylogeny of the Delphinidae. *Mol. Phylogenet. Evol.* **46**:252–268.
6. Casey, J. W., A. Roach, J. I. Mullins, K. B. Burck, M. O. Nicolson, M. B. Gardner, and N. Davidson. 1981. The U3 portion of feline leukemia virus DNA identifies horizontally acquired proviruses in leukemic cats. *Proc. Natl. Acad. Sci. USA* **78**:7778–7782.
7. Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**:3497–3500.
8. Coffin, J. M., S. H. Hughes, and H. E. Varmus. 1997. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
9. Delassus, S., P. Sonigo, and S. Wain-Hobson. 1989. Genetic organization of gibbon ape leukemia virus. *Virology* **173**:205–213.
10. Demirov, D. G., and E. O. Freed. 2004. Retrovirus budding. *Virus Res.* **106**:87–102.
11. Dickerson, F., E. Rubalcaba, R. Viscidi, S. Yang, C. Stallings, A. Sullens, A. Origeni, F. Leister, and R. Yolken. 2008. Polymorphisms in human endogenous retrovirus K-18 and risk of type 2 diabetes in individuals with schizophrenia. *Schizophr. Res.* **104**:121–126.
12. Dolei, A., and H. Perron. 2009. The multiple sclerosis-associated retrovirus and its HERV-W endogenous family: a biological interface between virology, genetics, and immunology in human physiology and disease. *J. Neurovirol.* **15**:4–13.
13. Donahue, P. R., E. A. Hoover, G. A. Beltz, N. Riedel, V. M. Hirsch, J. Overbaugh, and J. I. Mullins. 1988. Strong sequence conservation among horizontally transmissible, minimally pathogenic feline leukemia viruses. *J. Virol.* **62**:722–731.
14. Gifford, R., P. Kabat, J. Martin, C. Lynch, and M. Tristem. 2005. Evolution and distribution of class II-related endogenous retroviruses. *J. Virol.* **79**:6478–6486.
15. Hanger, J. J., L. D. Bromham, J. J. McKee, T. M. O'Brien, and W. F. Robinson. 2000. The nucleotide sequence of koala (*Phascolarctos cinereus*) retrovirus: a novel type C endogenous virus related to gibbon ape leukemia virus. *J. Virol.* **74**:4264–4272.
16. 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.
17. Ikeda, H., and H. Sugimura. 1989. Fv-4 resistance gene: a truncated endogenous murine leukemia virus with ecotropic interference properties. *J. Virol.* **63**:5405–5412.
18. Katsoulos, P. D., G. Christodouloupoloulos, G. Kontopidis, A. Minas, A. Tziava, and S. K. Kritas. 2009. Leukocyte counts in bronchoalveolar lavage fluid obtained from normal and maedi-visna-infected sheep. *Vet. Clin. Pathol.* **38**:397–402.
19. Kennedy-Stoskopf, S., M. K. Stoskopf, M. A. Eckhaus, and J. D. Strandberg. 1986. Isolation of a retrovirus and a herpesvirus from a captive California sea lion. *J. Wildl. Dis.* **22**:156–164.
20. Krach, U., N. Fischer, F. Czaderna, and R. R. Tonjes. 2001. Comparison of replication-competent molecular clones of porcine endogenous retrovirus class A and class B derived from pig and human cells. *J. Virol.* **75**:5465–5472.
21. Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief. Bioinform.* **5**:150–163.
22. Lapuk, A. V., P. P. Khil, I. V. Lavrentieva, Y. B. Lebedev, and E. D. Sverdlov. 1999. A human endogenous retrovirus-like (HERV) LTR formed more than 10 million years ago due to an insertion of HERV-H LTR into the 5' LTR of HERV-K is situated on human chromosomes 10, 19 and Y. *J. Gen. Virol.* **80**:835–839.
23. LeDuc, R. G., W. F. Perrin, and A. E. Dizon. 1999. Phylogenetic relationships among the delphinid cetaceans based on full cytochrome b sequences. *Mar. Mammal Sci.* **15**:619–648.
24. Ma, Y., M. Lu, S. Xu, J. Wu, K. Tian, and J. Zhang. 13 December 2008. Identification of full-length proviral DNA of porcine endogenous retrovirus from Chinese Wuzhishan miniature pigs inbred. *Comp. Immunol. Microbiol. Infect. Dis.* doi:10.1016/j.cimid.2008.10.007.
25. Marguerat, S., W. Y. Wang, J. A. Todd, and B. Conrad. 2004. Association of human endogenous retrovirus K-18 polymorphisms with type 1 diabetes. *Diabetes* **53**:852–854.
26. Martin, J., E. Herniou, J. Cook, R. W. O'Neill, and M. Tristem. 1999. Interclass transmission and phyletic host tracking in murine leukemia virus-related retroviruses. *J. Virol.* **73**:2442–2449.
27. Martin, J., E. Herniou, J. Cook, R. Waugh O'Neill, and M. Tristem. 1997. Human endogenous retrovirus type I-related viruses have an apparently widespread distribution within vertebrates. *J. Virol.* **71**:437–443.
28. Martineau, D., P. R. Bowser, R. R. Renshaw, and J. W. Casey. 1992. Molecular characterization of a unique retrovirus associated with a fish tumor. *J. Virol.* **66**:596–599.
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. Organisms* **10**:153–158.
30. Masuda, M., M. P. Remington, P. M. Hoffman, and S. K. Ruscetti. 1992. Molecular characterization of a neuropathogenic and nonerythroleukemogenic variant of Friend murine leukemia virus PVC-211. *J. Virol.* **66**:2798–2806.
31. Newman, S. J., and S. A. Smith. 2006. Marine mammal neoplasia: a review. *Vet. Pathol.* **43**:865–880.
32. Niebert, M., and R. R. Tonjes. 2005. Evolutionary spread and recombination of porcine endogenous retroviruses in the *Suiformes*. *J. Virol.* **79**:649–654.
33. Nikaïdo, M., F. Matsuno, H. Hamilton, R. L. Brownell, Jr., Y. Cao, W. Ding, Z. Zuoyan, A. M. Shedlock, R. E. Fordyce, M. Hasegawa, and N. Okada. 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* **98**:7384–7389.
34. Oliveira, N. M., H. Satija, I. A. Kouwenhoven, and M. V. Eiden. 2007. Changes in viral protein function that accompany retroviral endogenization. *Proc. Natl. Acad. Sci. USA* **104**:17506–17511.
35. Paul, T. A., S. L. Quackenbush, C. Sutton, R. N. Casey, P. R. Bowser, and J. W. Casey. 2006. Identification and characterization of an exogenous retrovirus from Atlantic salmon swim bladder sarcomas. *J. Virol.* **80**:2941–2948.
36. Perron, H., J. A. Garson, F. Bedin, F. Beseme, G. Paranhos-Baccala, F. Komurian-Pradel, F. Mallet, P. W. Tuke, C. Voisset, J. L. Blond, B. Lalande, J. M. Seigneurin, B. Mandrand, and the Collaborative Research Group on Multiple Sclerosis. 1997. Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci. USA* **94**:7583–7588.
37. Rastogi, P. A. 2000. MacVector. Integrated sequence analysis for the Macintosh. *Methods Mol. Biol.* **132**:47–69.
38. Reyes, R. A., and G. L. Cockerell. 1996. Unintegrated bovine leukemia virus DNA: association with viral expression and disease. *J. Virol.* **70**:4961–4965.
39. Roca, A. L., J. Pecon-Slatery, and S. J. O'Brien. 2004. Genomically intact endogenous feline leukemia viruses of recent origin. *J. Virol.* **78**:4370–4375.
40. Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
41. Ruprecht, K., F. Gronen, M. Sauter, B. Best, P. Rieckmann, and N. Mueller-Lantzsch. 2008. Lack of immune responses against multiple sclerosis-associated retrovirus/human endogenous retrovirus W in patients with multiple sclerosis. *J. Neurovirol.* **14**:143–151.
42. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
43. Sarkar, S., O. W. Jones, and N. Shioura. 1974. Constancy in human sperm DNA content. *Proc. Natl. Acad. Sci. USA* **71**:3512–3516.
44. Sheets, R. L., R. Pandey, W. C. Jen, and P. Roy-Burman. 1993. Recombinant feline leukemia virus genes detected in naturally occurring feline lymphosarcomas. *J. Virol.* **67**:3118–3125.
45. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature* **293**:543–548.
46. Stoye, J. P. 2006. Koala retrovirus: a genome invasion in real time. *Genome Biol.* **7**:241.
47. Tandon, R., V. Cattori, A. C. Pepin, B. Riond, M. L. Meli, M. McDonald, M. G. Doherr, H. Lutz, and R. Hofmann-Lehmann. 2008. Association between endogenous feline leukemia virus loads and exogenous feline leukemia virus infection in domestic cats. *Virus Res.* **135**:136–143.
48. Tandon, R., V. Cattori, B. Willi, M. L. Meli, M. A. Gomes-Keller, H. Lutz, and R. Hofmann-Lehmann. 2007. Copy number polymorphism of endogenous feline leukemia virus-like sequences. *Mol. Cell. Probes* **21**:257–266.
49. Tarlinton, R., J. Meers, J. Hanger, and P. Young. 2005. Real-time reverse transcriptase PCR for the endogenous koala retrovirus reveals an association between plasma viral load and neoplastic disease in koalas. *J. Gen. Virol.* **86**:783–787.
50. Tarlinton, R., J. Meers, and P. Young. 2008. Biology and evolution of the endogenous koala retrovirus. *Cell. Mol. Life Sci.* **65**:3413–3421.
51. Tarlinton, R. E., J. Meers, and P. R. Young. 2006. Retroviral invasion of the koala genome. *Nature* **442**:79–81.
52. Ting, Y. T., C. A. Wilson, K. B. Farrell, G. J. Chaudry, and M. V. Eiden. 1998. Simian sarcoma-associated virus fails to infect Chinese hamster cells despite the presence of functional gibbon ape leukemia virus receptors. *J. Virol.* **72**:9453–9458.
53. Tonjes, R. R., and M. Niebert. 2003. Relative age of proviral porcine endogenous retrovirus sequences in *Sus scrofa* based on the molecular clock hypothesis. *J. Virol.* **77**:12363–12368.
54. Tristem, M., P. Kabat, L. Lieberman, S. Linde, A. Karpas, and F. Hill. 1996. Characterization of a novel murine leukemia virus-related subgroup within mammals. *J. Virol.* **70**:8241–8246.
55. Urisman, A., R. J. Molinaro, N. Fischer, S. J. Plummer, G. Casey, E. A. Klein, K. Malathi, C. Magi-Galluzzi, R. R. Tubbs, D. Ganem, R. H. Silverman, and J. L. DeRisi. 2006. Identification of a novel gammaretrovirus in

- prostate tumors of patients homozygous for R462Q *RNAseL* variant. *PLoS Pathog.* **2**:e25.
56. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. *Science* **216**:812–820.
57. **Voisset, C., R. A. Weiss, and D. J. Griffiths.** 2008. Human RNA “rumor” viruses: the search for novel human retroviruses in chronic disease. *Microbiol. Mol. Biol. Rev.* **72**:157–196.
58. **Weiss, R. A.** 2006. The discovery of endogenous retroviruses. *Retrovirology* **3**:67.
59. **Wolgamot, G., L. Bonham, and A. D. Miller.** 1998. Sequence analysis of *Mus dunni* endogenous virus reveals a hybrid VL30/gibbon ape leukemia virus-like structure and a distinct envelope. *J. Virol.* **72**:7459–7466.
60. **Yonezawa, M., H. Nakamine, T. Tanaka, and T. Miyaji.** 1989. Hodgkin’s disease in a killer whale (*Orcinus orca*). *J. Comp. Pathol.* **100**:203–207.
61. **Zsiros, J., M. F. Jebbink, V. V. Lukashov, P. A. Voute, and B. Berkhout.** 1998. Evolutionary relationships within a subgroup of HERV-K-related human endogenous retroviruses. *J. Gen. Virol.* **79**:61–70.