Distribution of Endogenous Retroviruses in Crocodilians

Weerachai Jaratlerdsiri,1 Clara J. Rodriguez-Zárate,1 Sally R. Isberg,1,2 Chandramaya Siska Damayanti,1 Lee G. Miles,1 Nantarika Chansue,3 Chris Moran,1 Lorna Melville,4 and Jaime Gongora1*

Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, RMC Gunn Building, University of Sydney, Sydney, New South Wales 2006, Australia; Porosus Pty. Ltd., P.O. Box 86, Palmerston, Northern Territory 0831, Australia; Faculty of Veterinary Science, Chulalongkorn University, Pathumwan, Bangkok, 10330, Thailand; and OIC Berrimah Veterinary Laboratories, Department of Primary Industry, Fisheries and Mines, GPO Box 3000 Darwin, Northern Territory 0801, Australia

Received 1 April 2009/Accepted 6 July 2009

Knowledge of endogenous retroviruses (ERVs) in crocodilians (Crocodylia) is limited, and their distribution among extant species is unclear. Here we analyzed the phylogenetic relationships of these retroelements in 20 species of crocodilians by studying the pro-pol gene. The results showed that crocodilian ERVs (CERVs) cluster into two major clades (CERV 1 and CERV 2). CERV 1 clustered as a sister group of the genus Gammaretrovirus, while CERV 2 clustered distantly with respect to all known ERVs. Interestingly, CERV 1 was found only in crocodiles (Crocodyliidae). The data generated here could assist future studies aimed at identifying orthologous and paralogous ERVs among crocodilians.

Endogenous retroviruses (ERVs) are copies or remnants of exogenous viruses derived from past infections of germ cells and subsequent integration into the host genome (7, 8). Most ERVs are defective, having accumulated random inactivating mutations, and are therefore not pathogenic. However, many intact ERVs have been associated with different host diseases (4, 21). Those from pigs are considered potentially hazardous during xenotransplantation due to reactivation or recombination with exogenous retroviruses (18). ERVs have been studied extensively in mammals and birds (9, 10, 16), while knowledge of ERVs in reptiles is limited to a few host species (17, 24). Studies of a few full-length and partial pol gene sequences of some crocodilian species (order Crocodylia) have shown that some ERV sequences clustered distantly from Spumavirus while others clustered closely with Gammaretrovirus-like retroviruses (16, 17). However, their diversity, lineage specificity, and functionality have not been assessed across all extant crocodilian species.

The extant crocodylian lineage consists of 23 species classified in three families, the Alligatoridae (alligators), Crocodyliidae (crocodiles), and Gavialidae (gharials). The Crocodyliidae and Alligatoridae families are unambiguously recognized, with gharials either lumped with the Crocodylidae family or assigned in a separate family, the Gavialidae (6, 12, 19). It has been estimated, based on DNA and amino acid data, that the Alligatoridae and Crocodyliidae lineages diverged from a common ancestor about 97 to 103 million years ago (12, 19).

Here we analyze the distribution, potential to function, and phylogenetic relationships of ERVs in 20 extant crocodylian species by studying the protease-reverse transcriptase (pro-pol) gene. The pro-pol gene (0.8 to 1 kb) was amplified using two sets of degenerate primers (23). The PCR amplicons were gel purified and cloned using standard protocols, and about three ERV inserts were sequenced from each species by the sequencing service at the Australian Genome Research Facility. Sixty-five ERV DNA sequences were generated and translated to putative amino acid sequences using universal translation codes in the Molecular Toolkit interface (http://arbl.cvmbs.colostate.edu/molkit/index.html). Open reading frames and levels of similarity to other available ERV sequences were determined using the blastx tool available through NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Two aligned datasets were generated using the CLUSTALW program (22); the first consisted of 286 predicted amino acids from 65 novel and 9 published (16, 17) crocodilian ERVs (CERVs), and the second comprised 258 predicted amino acids from the novel crocodilian sequences and 73 published endogenous and exogenous retroviral sequences after exclusion of gaps. These known retroviruses included avian leukosis virus (Alpharetrovirus); murine leukemia virus (Gammaretrovirus); mouse mammary tumor virus and Jaagsiekte sheep retrovirus (Betaretrovirus); bovine leukemia virus and human T-lymphotropic virus (Delta-retrovirus); human immunodeficiency virus type 1, equine infectious anemia virus, and visna virus (Lentivirus); human foamy virus (Spumavirus); walleye dermal sarcoma virus (Epsilonretrovirus); 4 avian, 3 reptilian, 2 amphibian, and 14 mammalian ERVs, including human ERVs (13, 14); 3 chicken ERVs (10); 6 avian viruses similar to the alpha/beta group (7); and 22 Gammaretrovirus-like ERVs (9, 16). The best-fit model (JTT matrix model with parameter α = 1.61) for the two data sets was selected by using ProTest software (1) to implement neighbor-joining (NJ) in the MEGA 4 program (20). In addition, a maximum parsimony analysis was performed. Levels of amino acid similarity between CERV and other ERV sequences were also ascertained in the MEGA 4 program, using the p-distance option.
ERVs were found in all crocodilian lineages examined (Fig. 1). All 65 sequences show deletions, and 28 of these contain in-frame stop codons (Fig. 1). These mutations indicate that all CERVs generated in this study are defective and, therefore, nonfunctional, as has been observed in other vertebrate hosts (2, 3, 11, 15). Although only nonfunctional sequences were identified in members of the Alligatoridae and Crocodylidae, our study does not exclude the possibility of functional ERVs in these lineages.

The NJ and maximum parsimony phylogenetic trees were consistent and showed that CERVs cluster into two distinct major clades (Fig. 1), named CERV 1 and 2. CERV 1 consists of 44 sequences from 12 species of Crocodylidae, revealing for the first time the existence of a host lineage-specific ERV clade for crocodiles. CERV 2 consists of 30 sequences representing eight Alligatoridae species and the only Gavialidae species. Four sequences from three Crocodylidae species (Crocodylus niloticus, Crocodylus palustris, and Mecistops cataphractus) also clustered within CERV 2. Interestingly, sequences from both CERV 1 and 2 were found in a single Crocodylidae species (Crocodylus niloticus). Pairwise genetic distances show that the variation within CERV 1 (distance = 0.070 ± 0.006 [mean ± standard deviation]) is very much lower than that within CERV 2 (distance = 0.459 ± 0.020). Analyses also show that all CERV 2 sequences are distinct, revealing additional diversity and new minor clades within this unusual clade (Fig. 1) that have not been reported previously. In contrast, CERV 1 sequences show a high degree of similarity between species, and some of them are identical within species, including those from C. intermedius, C. siamensis, and C. moreletii.
FIG. 2. NJ analyses of 258 putative amino acids of the pol gene from CERVs and representatives from seven known ERV genera. Symbols represent the ERV hosts used in this study, as indicated in the key. Dashed circles show the two major CERV clades, CERV 1 and 2. Clustering information within CERVs is provided in Fig. 1. ALV, avian leukemia virus; BaEV, baboon endogenous retrovirus; BLV, bovine leukemia virus; Devl, Dendrobates Ventrimaculatus ERVI; EIAV, equine infectious anemia virus; ERV-3, human ERV3; GGERV, chicken ERV; HERV, human ERV; HFV, human foamy virus; HIV-1, human immunodeficiency virus type 1; HTLV, human T-lymphotropic virus; JSRV, Jaagsiekte sheep retrovirus; MMTV, mouse mammary tumor virus; MuLV, murine leukemia virus; PERV, porcine ERV; Spev, Sphenodon ERV; WDSV, walleye dermal sarcoma virus; Xen1, Xenopus laevis ERV1.
TABLE 1. Percentages of similarity among CERV 1 and CERV 2 and exogeneous members of the Retroviridae

<table>
<thead>
<tr>
<th>CERV clade</th>
<th>ALV</th>
<th>MMTV</th>
<th>MuLV</th>
<th>BLV</th>
<th>HIV-1</th>
<th>HFV</th>
<th>WDSV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>38</td>
<td>30</td>
<td>47</td>
<td>37</td>
<td>20</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>21</td>
<td>24</td>
<td>28</td>
<td>20</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

* Calculated from alignments of 265 amino acids of the pro-pol gene. Values are based on average similarity observed in CERV 1 and CERV 2 (Pairwise deletion option in the MEGA 4 program). ALV, avian leukaemia virus; MMTV, mouse mammary tumour virus; MuLV, murine leukaemia virus; BLV, bovine leukaemia virus; HIV-1, human immunodeficiency virus type 1; HFV, human foamy virus; WDSV, wallaby dermal sarcoma virus.

Analysis of phylogenetic relationships with known ERVs showed that CERV 1 clusters closely with two clades of Gammaretrovirus (birds and reptiles/mammals/amphibians/human), showing polytomy and long branch lengths with respect to each other (Fig. 2). Cade CERV 2 clustered distantly from all known ERVs (Fig. 2). These relationships are also supported by the amino acid similarity between a known representative of Gammaretrovirus, murine leukemia virus, and CERV clades. In agreement with the results of previous CERV studies (17), the amino acid similarity was lower for CERV 2 (24%) and higher for CERV 1 (47%) (Table 1).

The comparison of CERV and host species phylogenies (6, 12, 19) showed discordance. While crocodilian host phylogenies based on DNA sequence data are quite well defined, this was not the case within the CERV clades. Both CERV 1 and 2 showed a mixture of internal topologies from symmetric (bushlike), to random, to asymmetric with differential branch lengths, making it difficult to assess any coevolutionary patterns. Given that CERV 1 appeared to be specific to the Crocodyliidae family and that its internal branch lengths were considerably shorter than those observed in the crocodile host phylogeny (5) and CERV 2, it is plausible that CERV 1 has a lower mutation rate and represents a relatively recent retroviral infection that occurred after the divergence of the Crocodylia and Alligatoridae families from the common ancestor.

The current investigation has confirmed the existence of two groups of ERVs and revealed additional distribution and diversity among extant members of the Crocodylia. Interestingly, we found a host lineage-specific clade which could have potential for use in the identification of members of the Crocodylia at the family level. The data generated here will assist future studies identifying orthologous and paralogous ERVs among crocodilian species to assess the variation, distribution, and taxonomy of these retroelements within crocodilian species and populations.

Nucleotide sequence accession numbers. GenBank accession numbers for sequences derived in this study are FJ155497 through FJ155561.

We thank Travis Glenn, Kent Vliet, Robert Godshalk, Mitch Eaton, and Matthew Shirley who kindly provided us with many of the crocodilian DNA samples included in this investigation. We are also grateful to Porosus Pty. Ltd. for providing the Australian saltwater and Johnston’s crocodile samples and to Panya Younggrapaporn of Golden Crocodile Agriculture Co. Ltd., Thailand, and Thanida Haetrakul of Chulalongkorn University, Thailand, for providing us with samples from Siamese crocodile species.

REFERENCES