



SHORT REPORT

Open Access

Extensive retroviral diversity in shark

Guan-Zhu Han^{1,2}

Abstract

Background: Retroviruses infect a wide range of vertebrates. However, little is known about the diversity of retroviruses in basal vertebrates. Endogenous retrovirus (ERV) provides a valuable resource to study the ecology and evolution of retrovirus.

Findings: I performed a genome-scale screening for ERVs in the elephant shark (*Callorhynchus milii*) and identified three complete or nearly complete ERVs and many short ERV fragments. I designate these retroviral elements “*C. milli* ERVs” (CmiERVs). Phylogenetic analysis shows that the CmiERVs form three distinct lineages. The genome invasions by these retroviruses are estimated to take place more than 50 million years ago.

Conclusions: My results reveal the extensive retroviral diversity in the elephant shark. Diverse retroviruses appear to have been associated with cartilaginous fishes for millions of years. These findings have important implications in understanding the diversity and evolution of retroviruses.

Keywords: Endogenous retroviruses, Chondrichthyes, Paleovirology

Findings

Retroviruses infect a wide range of vertebrates and cause many notorious diseases, such as AIDS and cancers. However, much remains unknown about the diversity of retroviruses in basal vertebrate species. In particular, only several retroviruses have been identified in fishes, including Snakehead retrovirus, walleye dermal sarcoma virus, walleye epidermal hyperplasia virus, and Atlantic salmon swim bladder sarcoma virus [1-4]. Retrovirus employs a unique replication strategy, which requires reverse transcription of its RNA genome into DNA and integration of viral DNA into the host chromosomes. Occasionally, retroviruses infect germ line cells, and the resulting integrated retrovirus, known as endogenous retrovirus (ERV), becomes vertically inherited as a host genomic locus. Over time, some retroviral insertions are fixed in the host population. ERVs provide important insights into the ecology and evolutionary history of retroviruses.

Cartilaginous fishes (Chondrichthyes) are the most basal class of vertebrates from which retrovirus has been

reported [5]. Here, I analyzed the recently available genome sequence of the elephant shark (*Callorhynchus milii*), a high-quality genome assembly covering approximately 94% of the *C. milii* genome, for retroviral insertions [6]. The tBLASTn algorithm with various representative retroviral Pol protein sequences was employed to screen the elephant shark genome for candidate ERV sequences. To distinguish ERVs from other LTR-retrotransposons, I used a strict criterion: only the retroviral Pol protein homolog sequence with a downstream Env protein homolog is defined as an ERV element. After initial identification of ERVs, the BLASTn algorithm was used to identify short ERV fragments. My genome-scale screening procedure identified three complete or nearly complete ERV insertions (within the *C. milii* genome scaffolds 2, 324, and 2324, respectively; Additional file 1: Dataset 1) and many short ERV fragments in the elephant shark genome. I designate these retroviral elements “*C. milli* ERVs” (CmiERVs).

To assess the relationship between CmiERVs and other retroviruses, CmiERV and representative retroviral Pol protein sequences (Additional file 2: Table S1) were aligned using MUSCLE [7]. The ambiguous regions in the Pol protein alignments were removed using Gblocks 0.91b and then manually edited [8]. Phylogenetic analyses were performed using MrBayes 3.1.2 [9]. My phylogenetic analysis shows that these CmiERVs form three distinct lineages (Figure 1). Lineage I CmiERVs cluster

Correspondence: guan Zhu@email.arizona.edu

¹Jiangsu Key Laboratory for Microbes and Functional Genomics, Jiangsu Engineering and Technology Research Center for Microbiology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu 210023, China

²Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85719, USA

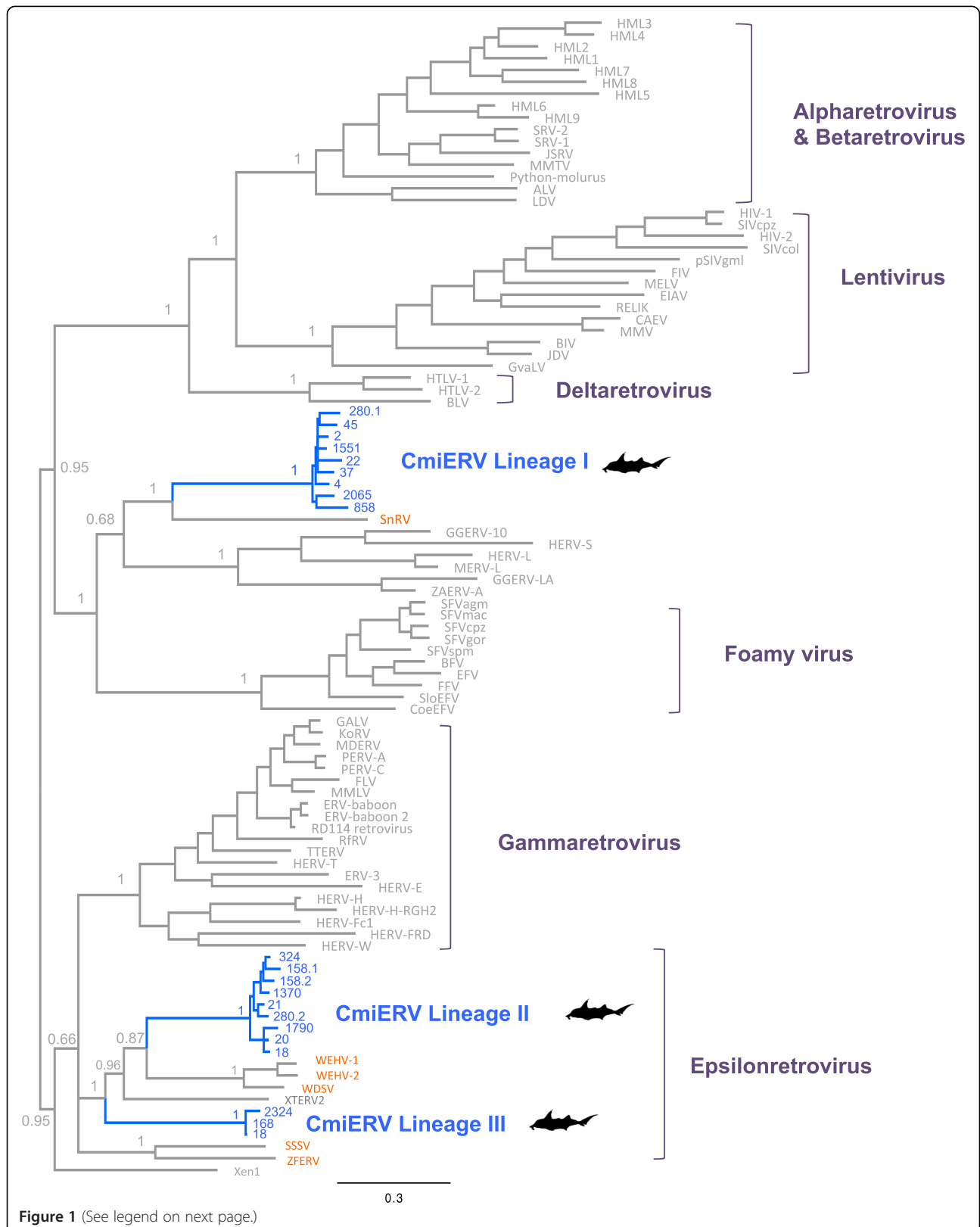


Figure 1 (See legend on next page.)

(See figure on previous page.)

Figure 1 Phylogeny of CmiERVs and other representative retroviruses. The phylogeny was reconstructed based on the retrovirus Pol protein sequences. Posterior probabilities are labeled near the selected nodes. The CmiERVs and the retroviruses of fish origin are highlighted in blue and orange, respectively. The numbers of the scaffolds where CmiERV fragments were identified are labeled near the corresponding tips.

Abbreviation: ALV, Avian leukosis virus; BFV, Bovine foamy virus; BIV, Bovine immunodeficiency virus; BLV, Bovine leukemia virus; CAEV, Caprine arthritis-encephalitis virus; CoEFV, Coelacanth endogenous foamy virus; EFV, Equine foamy virus; EIAV, Equine infectious anemia virus; ERV-baboon, Baboon endogenous virus; FFV, Feline foamy virus; FIV, Feline immunodeficiency virus; FLV, Feline leukemia virus; GALV, Gibbon ape leukemia virus; GGERV, *Gallus gallus* endogenous retrovirus; GvaELV, *Galeopterus variegatus* endogenous lentivirus; HIV-1, Human immunodeficiency virus type 1; HIV-2, Human immunodeficiency virus type 2; HML1-9, Human MMTV-like 1–9; HTLV, Human T-cell leukemia virus; HERV, Human endogenous retrovirus; JDV, Jembrana disease virus; JSRV, Jaagsiekte sheep retrovirus; KoRV, Koala retrovirus; LDV, Lymphoproliferative disease virus; MDERV, *Mus dunni* endogenous retrovirus; MELV, *Mustelidae* endogenous Lentivirus; MERV-L, Murine endogenous retrovirus type L; MMLV, Moloney murine leukemia virus; MMTV, Mouse mammary tumor virus; MVV, Maedi-visna virus; PERV-A, Porcine endogenous retrovirus A; PERV-C, Porcine endogenous retrovirus C; pSIVgml, Gray mouse lemur prosimian immunodeficiency virus; Python-molurus, Python molurus endogenous retrovirus; RELIK, Rabbit endogenous lentivirus type K; RfRV, *Rhinolophus ferrumequinum* retrovirus; SFV, Simian foamy virus; SIV, Simian immunodeficiency virus; SloEFV, Sloth endogenous foamy virus; SnRV, Snakehead retrovirus; SRV-1, Simian retrovirus 1; SRV-2, Simian retrovirus 2; SSSV, Atlantic salmon swim bladder sarcoma virus; TTERV, *Tursiops truncatus* endogenous retrovirus; WDSV, Walleye dermal sarcoma virus; WEHV-1, Walleye epidermal hyperplasia virus type 1; WEHV-2, Walleye epidermal hyperplasia virus type 2; XTERV2, *Xenopus tropicalis* endogenous retrovirus 2; ZAERV-A, *Zonotrichia albicollis* endogenous retrovirus type A; ZFERV, Zebrafish endogenous retrovirus.

with the retroviruses isolated from the snakehead fish (*Ophicephalus striatus*), while lineage II and III CmiERVs cluster with the epsilonretroviruses isolated from the walleye (*Sander vitreus*) and amphibians. CmiERV lineage I and lineages II/III are only distantly related to each other. These CmiERV lineages are likely to result from three independent retroviral invasion events. My analysis provides clear evidence there is extensive retroviral diversity in the elephant shark. It is possible that exogenous retroviruses related to CmiERVs identified here are still circulating in the elephant shark and possibly other Chondichthyes.

On endogenization, the 5' LTR and 3' LTR of a nascent ERV are identical and will accumulate mutations independently. Thus, the 5' LTR and 3' LTR genetic divergence could be used to estimate ERV invasion time [10]. The invasion time of an ERV can be estimated through:

$$t = \frac{d}{2u}$$

where t indicates the invasion time, u indicates the neutral evolutionary rate of host, and d indicates the genetic divergence between 5' LTR and 3' LTR. In this study, two complete CmiERV insertions were identified (Table 1). The genetic divergence between 5'-LTR and 3'-LTR was calculated with the Kimura two-parameter substitution model [11]. The neutral evolutionary rate for the elephant shark is not available but is approximately an order of magnitude lower than those for mammals [12]. The average neutral rate for mammals is

estimated to be 2.2×10^{-9} substitutions per site per year [13]. I thus assume 2.2×10^{-10} substitutions per site per year as the shark neutral rate. Then the insertion times for the two complete ERVs are approximately 75 and 54.5 million years ago, respectively. However, these estimates should be taken with cautions [14], given that I am not sure whether LTRs evolve at a neutral manner and what the actual neutral rate for the elephant shark is. Nevertheless, these results suggest these retroviruses were infecting the elephant shark millions of years ago.

Previously, a single ERV sequence was identified in the lemon shark (*Negaprion brevirostris*), which is closely related to human ERV; this ERV was thought to have a cross-transmission origin [5]. However, I find that CmiERVs cluster together with retroviruses of fish origin. The phylogenetic pattern is compatible with the hypothesis of an ancient marine origin of retroviruses [15]. Chondrichthyes are the most basal class of vertebrates from which retrovirus has been identified; no retrovirus is identified in earlier-diverging vertebrate lineages, the lampreys (Cephalaspidomorpha) and the hagfish (Myxini) [5]. It follows that these CmiERV elements are likely to represent "primitive" retroviruses. However, the possibility that these elephant shark retroviruses originated from cross-transmission from other fishes cannot be formally excluded.

To date, only a limited number of exogenous/endogenous retroviruses have been identified in fishes [1-5,14]. My results reveal the unexpectedly extensive retroviral diversity of the elephant shark. The initial candidate ERVs were identified based on a strict criterion – whether there is a downstream Env protein homolog following the Pol protein

Table 1 Genomic position and invasion time of two complete CmiERV insertions

Complete CmiERV	Lineage	Genomic position	5'- and 3'LTR divergence	Time (MYA)
1	I	Scaffold 2: 5,443,211-5,452,770	0.033	75
2	II	Scaffold 324: 74,676-84,761	0.024	54.5

homolog. This approach is conservative, given that the retroviral Env protein evolves rapidly and its similarity to other retroviral Env proteins will erode over a long time. On the other hand, the ERVs identified using this approach are authentic retroviruses. It is likely that there are additional ERV insertions that were not detected. Also, it should be noted that only a small proportion of retroviruses could leave endogenous copies in their host genomes [16]. Therefore, I believe the actual diversity of retroviruses is more extensive in the elephant shark. Further analysis of ERV in basal vertebrates would improve our understanding of the diversity and evolution of retroviruses.

Additional files

Additional file 1: Dataset 1. Sequences of three complete or nearly complete CmiERVs.

Additional file 2: Table S1. The representative retrovirus sequences used for phylogenetic reconstruction.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

GZH designed and performed the research, analyzed the data, and wrote the manuscript.

Acknowledgments

This research was supported by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions. I thank Chunyang Wang for helping prepare Figure 1.

Received: 12 December 2014 Accepted: 23 March 2015

Published online: 28 April 2015

References

- Hart D, Frerichs GN, Rambaut A, Onions DE. Complete nucleotide sequence and transcriptional analysis of snakehead fish retrovirus. *J Virol*. 1996;70:3606–16.
- LaPierre LA, Holzschu DL, Bowser PR, Casey JW. Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: evidence for a gene duplication. *J Virol*. 1999;73:9393–403.
- Holzschu DL, Martineau D, Fodor SK, Vogt VM, Bowser PR, Casey JW. Nucleotide sequence and protein analysis of a complex piscine retrovirus, walleye dermal sarcoma virus. *J Virol*. 1995;69:5320–31.
- Paul TA, Quackenbush SL, Sutton C, Casey RN, Bowser PR, Casey JW. Identification and characterization of an exogenous retrovirus from atlantic salmon swim bladder sarcomas. *J Virol*. 2006;80:2941–8.
- Herniou E, Martin J, Miller K, Cook J, Wilkinson M, Tristem M. Retroviral diversity and distribution in vertebrates. *J Virol*. 1998;72:5955–66.
- Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, et al. Elephant shark genome provides unique insights into gnathostome evolution. *Nature*. 2014;505:174–9.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
- Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol*. 2007;56:564–77.
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.
- Johnson WE, Coffin JM. Constructing primate phylogenies from ancient retrovirus sequences. *Proc Natl Acad Sci USA*. 1999;96:10254–60.
- Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol*. 1980;16:111–20.
- Martin AP. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). *Mol Biol Evol*. 1999;16:996–1002.
- Kumar S, Subramanian S. Mutation rates in mammalian genomes. *Proc Proc Natl Acad Sci USA*. 2002;99:803–8.
- Han GZ, Worobey M. A primitive endogenous lentivirus in a colugo: insights into the early evolution of lentiviruses. *Mol Biol Evol*. 2015;32:211–5.
- Han GZ, Worobey M. An endogenous foamy-like viral element in the coelacanth genome. *PLoS Pathog*. 2012;8:e1002790.
- Emerman M, Malik HS. Paleovirology—modern consequences of ancient viruses. *PLoS Biol*. 2010;8:e1000301.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

