Virus Research 262 (2019) 10–14

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Short communication

An endogenous adeno-associated virus element in elephants

Yuki Kobayashi[a,*], Tsukika Shimazu[a], Koichi Murata[b], Takuya Itou[a], Yoshiyuki Suzuki[c]

[a] Nihon University Veterinary Research Center, Fujisawa, Kanagawa 252-0880, Japan
[b] Department of Animal Resource Science, College of Bioresource Sciences, Nihon University, 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan
[c] Graduate School of Natural Sciences, Nagoya City University, Nagoya, Aichi 467-8501, Japan

ARTICLE INFO

Keywords:
Adeno-associated virus
Endogenous viral element
Elephant
Rep gene

ABSTRACT

An endogenous viral element derived from adeno-associated virus containing a nearly intact open reading frame (ORF) of the *rep* gene (enAAV-*rep*) has been identified in the genomes of various mammals including degu and African elephant. Particularly, in degu, mRNA expression of enAAV-*rep* has been observed specifically in the liver. Here we newly identified enAAV-*rep* in Asian elephant and rock hyrax, both of which are afrotherians. The enAAV-*rep* of African and Asian elephants appeared to be orthologous and originated from an integration event of the entire genome of AAV into the ancestral genome of elephants more than 6 million years ago, whereas that of rock hyrax appeared to have originated independently. Negative selection operating at the amino acid sequence level was detected for the ORF of enAAV-*rep* in elephants. As in degu, mRNA expression of enAAV-*rep* was specifically observed in the liver in Asian elephant. Integrations of enAAV-*rep* appeared to have occurred independently on the evolutionary lineages of elephants and degu, suggesting that the AAV Rep protein has been co-opted repeatedly in the mammalian liver.

1. Introduction

Endogenous viral elements (EVEs) are the fossil record of ancient viruses that have infected germ-line cells in the ancestors of host species. Although EVEs derived from various viruses have been discovered in many eukaryotes (Feschotte and Gilbert, 2012; Katzourakis and Gifford, 2010; Horie et al., 2010, 2013, 2016), most of them do not appear to encode functional proteins because of the existence of premature termination codons in the open reading frames (ORFs) (Gifford and Tristem, 2003). However, functional EVEs have been occasionally identified (Aswad and Katzourakis, 2012; Fujino et al., 2014; Kobayashi et al., 2016; Myers et al., 2016). For example, EVEs derived from the envelope genes of retroviruses are known to be involved in the placentation of mammals (Rote et al., 2004).

EVEs derived from adeno-associated virus (enAAVs) have been discovered in the genomes of several animals (Kapoor et al., 2010; Katzourakis and Gifford, 2010; Arriagada and Gifford, 2014; François et al., 2016). Adeno-associated virus (AAV) is a member of the genus *Dependovirus* in the family *Parvoviridae*, and possesses a linear, single-stranded DNA (ssDNA) genome of approximately 4.7 kb long. The genome comprises untranslated regions (UTRs) located at both ends containing inverted terminal repeats (ITRs) and ORFs for the *rep* gene, which encodes for non-structural proteins, and the *cap* gene, which encodes for structural proteins (Srivastava et al., 1983; Henckaerts and Linden, 2010). AAV replicates in the nucleus of infected cells and integrates its entire genome into the host genome in a site-specific manner (Kotin et al., 1990, 1991). ITRs are required for the viral genome to be integrated into and rescued from the host genome (Linden et al., 1996; Ling et al., 2015; Wang et al., 1996).

It has been reported that the EVE derived from the *rep* gene of AAV (enAAV-*rep*) maintains a nearly intact ORF in the genomes of degu (*Octodon degus*), long-tailed chinchilla (*Chinchilla lanigera*), olive baboon (*Papio anubis*), and African elephant (*Loxodonta africana*) (Katzourakis and Gifford, 2010; Arriagada and Gifford, 2014). Particularly, in degu, mRNA expression of enAAV-*rep* was specifically observed in the liver, supporting its functionality as a protein in the host (Arriagada and Gifford, 2014). In this study, we focus on enAAV-*rep* in elephants, which are members of Afrotheria. We show evidence that the enAAV-*rep* is functional as a protein in elephants.

2. Results

The tBLASTn search was conducted against the whole genome shotgun database (wgs) of mammals (taxid: 40674) using the amino

Abbreviations: AAV, adeno-associated virus; BAAV, bovine AAV; enAAV-*rep*, endogenous adeno-associated virus element derived from the *rep* gene; EVEs, endogenous virus elements; ITRs, inverted terminal repeats; ORF, open reading frame; UTR, untranslated region

* Corresponding author at: 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan.
E-mail address: suzuki.yukikb@nihon-u.ac.jp (Y. Kobayashi).

https://doi.org/10.1016/j.virusres.2018.04.015
Received 9 March 2018; Received in revised form 17 April 2018; Accepted 23 April 2018
Available online 24 April 2018
0168-1702/

Table 1
enAAV-*rep* identified in the genomes of afrotherians by tBLASTn search.

Species	Accession number	Orientation	Scafold		BAAV genome[a]	
			Start	End	Start	End
Loxodonta africana	AAGU03013549	+	51509	53245	2	492
Procavia capensis	ABRQ02194032	−	49775	48384	2	467

[a] Accession number: YP_024970.

```
BAAV            1 MATFYEVIVRVPFDVEEHLPGISDNFVDWVTGQIWELPPESDLNLTLIEQPQLTVADRIRRVFLYEWNKFSKQESKFFVQFEKGSEYFHLHTLVETSGISSMVLGRYVSQ 110
African elephant 1 --S...F.PSL.T...D...SVP.S..T.LLEAEIL..DW...EFDQ..MHHV.LG.KLF.EMHKF.CGCL.K.......L...TDFT...C.L.AD.VK...V...LN. 108
Rock hyrax      1 -.C.*.I.I..LS.I.......P.SL..RI...E*..S....FDVSQT......Q.EKLQW*..N..TQIT.SDPSY.I.LI...Q.....M..SSE.MKY......LN. 109

                111 IRAQLVKVVFQNIEPRINDWVAITKVKKGGANKVVDSGYIPAYLLPKVQPELQWAWTNLEEYKLAALNLEERKRLVAQFQLESSQRSQEASSQRDVSADPVIKSKTSQKY 220
                109 .KEN.I.R..RG...KVPN.L.A..T.......LRSQS.........S..........I.G..K.T.T.S.LA..L.E-----HGA.RP..AGLQFENG....T.Q..N. 213
                110 .ATR.CSQI...L..GFG.*L.V..P.HN..S.L*.MN...TS....QK....*...D.G...S........R..LNL.LGAR.LD.P.SQDNQ.SNPP.LYC..GM.Q. 219

                221 MALVSWLVEHGITSEKQWIQENQESYLSFNSTGNSRSQIKAALDNASKIMSLTKSASDYLVGQTVPEDISENRIWQIFDLNGYDPAYAGSVLYGWCTRAFGKRNTVWLYG 330
                214 .E..N....N.............N......A.S.E.A...S....S..........P.....ADP...LTQ...Y.L.QM.N........I.L...E.S.N...A...F. 323
                220 ....D...KN.V..K.......KD....Y.ASPE.Q.*.QC....TC.......T.A...I---T.---------.RH.H*PHPPSRLLA...LR..Q.K...W..I---- 314

                331 PATTGKTNIAEAISHTVPFYGCVNWTNENFPFNDCVEKMLIWWEEGKMTSKVVEPAKAILGGSRVRVDQKCKSSVQVDSTPVIITSNTNMCVVVDGNSTTFEHQQPLEDR 440
                324 ........L....A..M.................N.FD.......A....A.........F...K........V.Q.I.A........RD..M...........K..... 433
                315 -.I.....M....V.A..L...I..P.K....S...D.VVV........-....S........WIL.G*N....I.IH..L..........C.I.........WR..... 422

                441 MFRFELMRRLPPDFGKITKQEVKDFFAWAKVNQVPVTHEFMVPKKVAGTERAETSRKRPLDDVTNTNYKSPEKRARLSVVPETPRSSDVPVEPAPLRPLNWSSRYECRCD 550
                434 ....VFSK..E........R...E..K.VEL.PREIR.K.I.KHSLSACVGSNSVKPA.C.K----G......CP....AA.I.IEV..-IHRRSGKISDYLT..P.P.E 538
                423 ..K.K.VVQ.E.N...V..EK.....C---------------------------------------------------------------------------------- 449

                551 YHAKFDSVTGECDECEYLNRGKNGCIFHNATHCQICHAVPPWEKENVSDFNDFDDCNKEQ- 610
                539 F..DSVKLLEP.V.......K....LD.GV.N.PK..GL---------------------- 577
                449 ------------------------------------------------------------ 449
```

Fig. 1. enAAV-*rep* identified in the genomes of afrotherians by tBLASTn search using the amino acid sequence of BAAV Rep protein (accession number: YP_024970). *Premature termination codon.

acid sequence of Rep protein (618 aa) for bovine AAV (BAAV) as the query (accession number: YP_024970) (Schmidt et al., 2004). With the e-value threshold set at 10^{-100}, 18 hits (enAAV-*rep*) were obtained from 16 mammals (Fig. 3). In African elephant, a single copy of enAAV-*rep* containing a nearly intact ORF encoding 587 aa was detected with the e-value of 0, consistent with the previous study (Table 1, Figs. 1, S1) (Kapoor et al., 2010; Katzourakis and Gifford, 2010; Arriagada and Gifford, 2014). The amino acid sequence encoded by African elephant enAAV-*rep* and that of BAAV Rep shared 61% identity. In addition, a single copy of enAAV-*rep* was newly identified with the e-value of 10^{-127} from rock hyrax (*Procavia capensis*), which is another member of Afrotheria. However, premature termination codons were observed in the ORF, suggesting that rock hyrax enAAV-*rep* does not encode a functional protein (Fig. 1).

It should be noted that the wgs database might contain sequences derived from contaminating or infecting exogenous AAV (Kryukov et al., 2018). To confirm that the African elephant enAAV-*rep* identified using tBLASTn search was a genuine EVE, we conducted PCR using the DNA collected from muscle and liver tissues of Asian elephant (*Elephas maximus*) as well as LACF-NaNaII cell line of African elephant (RIKEN Cell BANK: RCB2320). The primers used for PCR were designed based on the nucleotide sequence of African elephant enAAV-*rep* (Table S1). The amplicons with the expected size were detected in muscle and liver tissues of Asian elephant as well as in the LACF-NaNaII cell line of African elephant (Fig. 2). Asian elephant enAAV-*rep* contains an ORF encoding the amino acid sequence of 587 aa, which was the same size and 96.8% identical to that of African elephant enAAV-*rep* (accession number: LC372536) (Fig. S1).

The phylogenetic tree was constructed for 18 enAAV-*rep* sequences obtained from 16 mammals in tBLASTn together with AAV *rep* using the amino acid sequences encoded by the ORFs. enAAV-*rep* from African and Asian elephants clustered with 100% bootstrap probability (Fig. 3). Although the bootstrap probability was less than 70%, the cluster of elephant enAAV-*rep* further made a cluster with BAAV *rep*, which in turn clustered with other enAAV-*rep* and AAV *rep*. The topology was consistent with that reported in the previous paper (Arriagada and Gifford, 2014). Rock hyrax enAAV-*rep* was found to be distantly related to elephant enAAV-*rep*. The 350 nt upstream and downstream regions of enAAV-*rep* were alignable between African and Asian elephants, with similar distribution patterns of retrotransposons (Fig. 4), indicating that the enAAVs identified in the genomes of African and Asian elephants were genuine EVEs having orthologue relationships. By contrast, the same regions of rock hyrax enAAV-*rep* were not alignable to those of elephant enAAV-*rep*, with a distinct distribution pattern of retrotransposons.

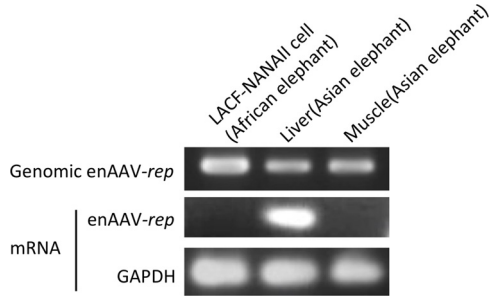

Fig. 2. Detection of genomic and mRNA enAAV-*rep* in elephant cell line and tissues. DNA and RNA were extracted from the LACF-NaNaII cell line of African elephant and liver and muscle tissues collected from a male Asian elephant that died of natural causes at Kobe Oji Zoo in Japan. Genomic enAAVs (486-bp product) were amplified by PCR (top panel). After DNA degradation in RNA samples using DNase, mRNA expression was investigated by RT-PCR. mRNAs transcribed from enAAV-*rep* were detected as a 207-bp product (middle panel). GAPDH mRNA was used as a positive control for mRNA detection (bottom panel). Primers used for the PCR and RT-PCR are shown in Table S1.

AAV is known to integrate its entire genome into the host genome in a site-specific manner (Kotin et al., 1990, 1991). Consequently, the EVE derived from the *cap* gene of AAV (enAAV-*cap*) has been observed in the vicinity of enAAV-*rep* in some mammalian genomes (Kapoor et al., 2010; Katzourakis and Gifford, 2010; Arriagada and Gifford, 2014). However, enAAV-*cap* was not identified in the genomes of elephants

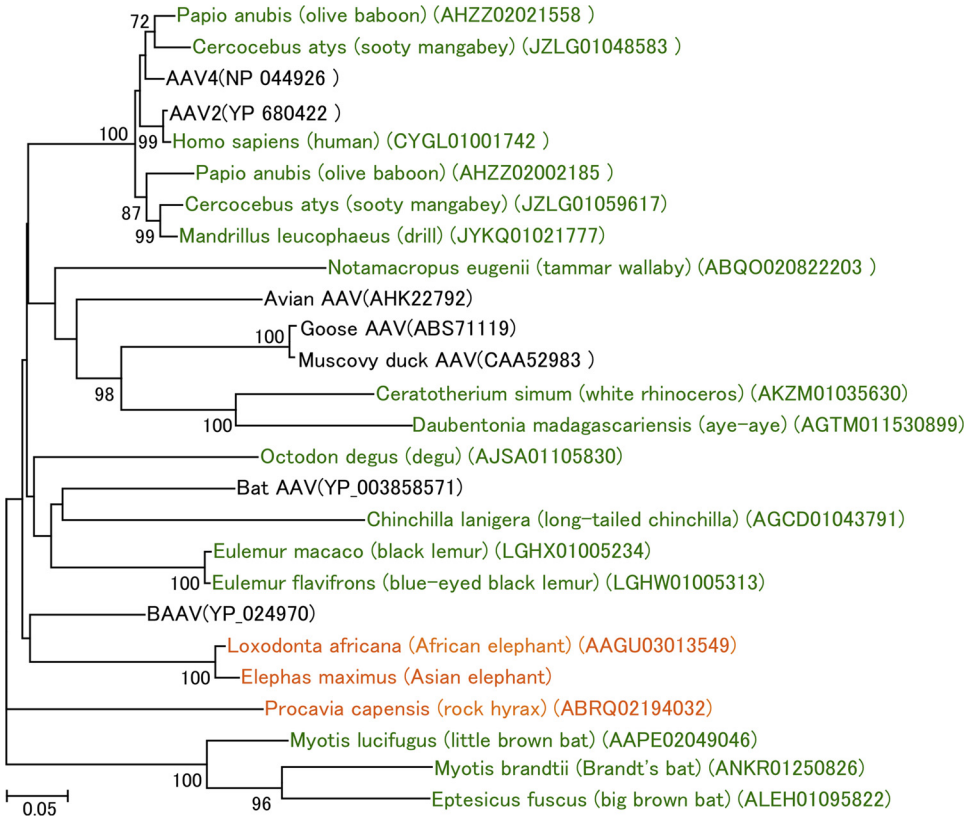

Fig. 3. Phylogenetic tree of enAAV-*rep* and exogenous AAV Rep protein. The phylogenetic tree was constructed by the neighbor-joining method with p-distance. The reliability of interior branches in the phylogenetic tree was assessed by computing the bootstrap probability with 1000 resamplings. Bootstrap values > 70% are shown for interior branches. The common names of hosts and accession numbers of nucleotide sequences containing enAAV-*rep* are shown in parentheses. Orange and green letters are used for describing enAAV-*rep* found in afrotherians and other mammals, respectively.

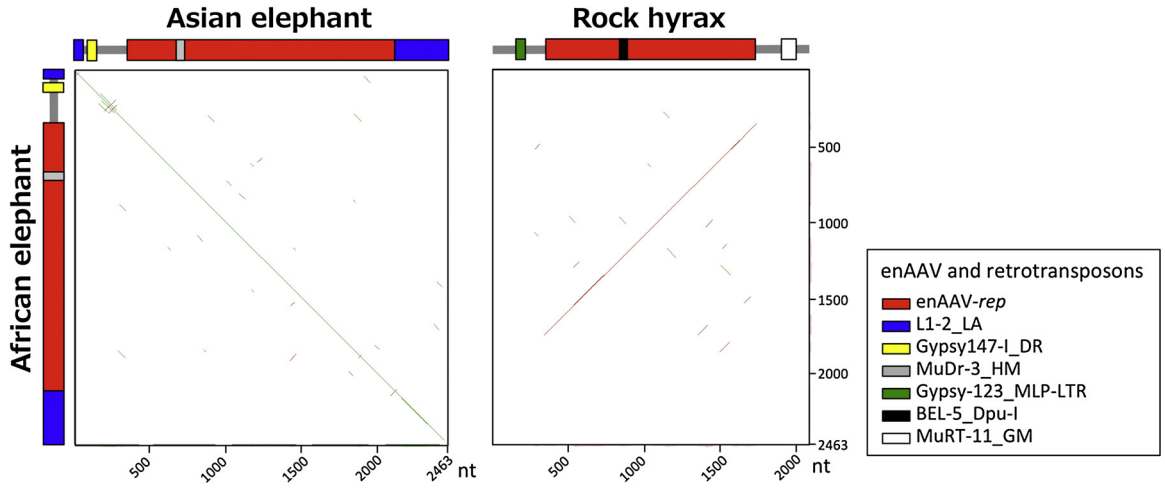

Fig. 4. Dot-plot analysis of enAAV-*rep* and its 350 nt upstream and downstream nucleotide sequences. Green and red lines in the boxes indicate forward- and reverse-match positions between compared sequences, respectively. Colored boxes on gray bars represent the positions of enAAV-*rep* and retrotransposons predicted by CENSOR software (Kohany et al., 2006).

and rock hyrax. Instead, the upstream ~150 nt region of elephant enAAV-*rep* aligned with the 5′ UTR of BAAV (Fig. 5), although the ITR structure was not predicted within the upstream 350 nt region of elephant enAAV-*rep* (data not shown). These results are consistent with the idea that enAAV-*rep* of African and Asian elephants are orthologous and have originated from an integration event of the entire genome of AAV into the ancestral genome of elephants more than 6 million years ago (Hedges et al., 2006). enAAV-*rep* may have originated independently on the lineage of rock hyrax after divergence from the lineage of elephants within afrotherians (Nishihara and Okada, 2008).

Since a nearly intact ORF appears to have been maintained in elephant enAAV-*rep*, it was hypothesized that the ORF of elephant enAAV-*rep* encodes a functional protein. To test this hypothesis, the d_S and d_N values were computed from the comparison of ORFs in enAAV-*rep* of African and Asian elephants using MEGA (version 7.0.21) (Kumar et al., 2016). The d_S and d_N values were estimated to be 0.0335 ± 0.009 and 0.0141 ± 0.003, respectively. Thus, the d_N/d_S ratio was 0.41 and negative selection was detected with $P < 0.05$ by the *Z*-test. Furthermore, mRNA expression of enAAV-*rep* was examined using the LACF-NaNaII cell line of African elephant and muscle and liver tissues of Asian elephant. mRNA expression was identified in the liver tissue of Asian elephant, but not in the LACF-NaNaII cell line of African elephant and the muscle tissue of Asian elephant (Fig. 2). Note that specific mRNA expression of enAAV-*rep* in the liver has also been observed in degu (Arriagada and Gifford, 2014). In the phylogenetic tree of enAAV-*rep* and AAV *rep* sequences, degu enAAV-*rep* was distantly related to

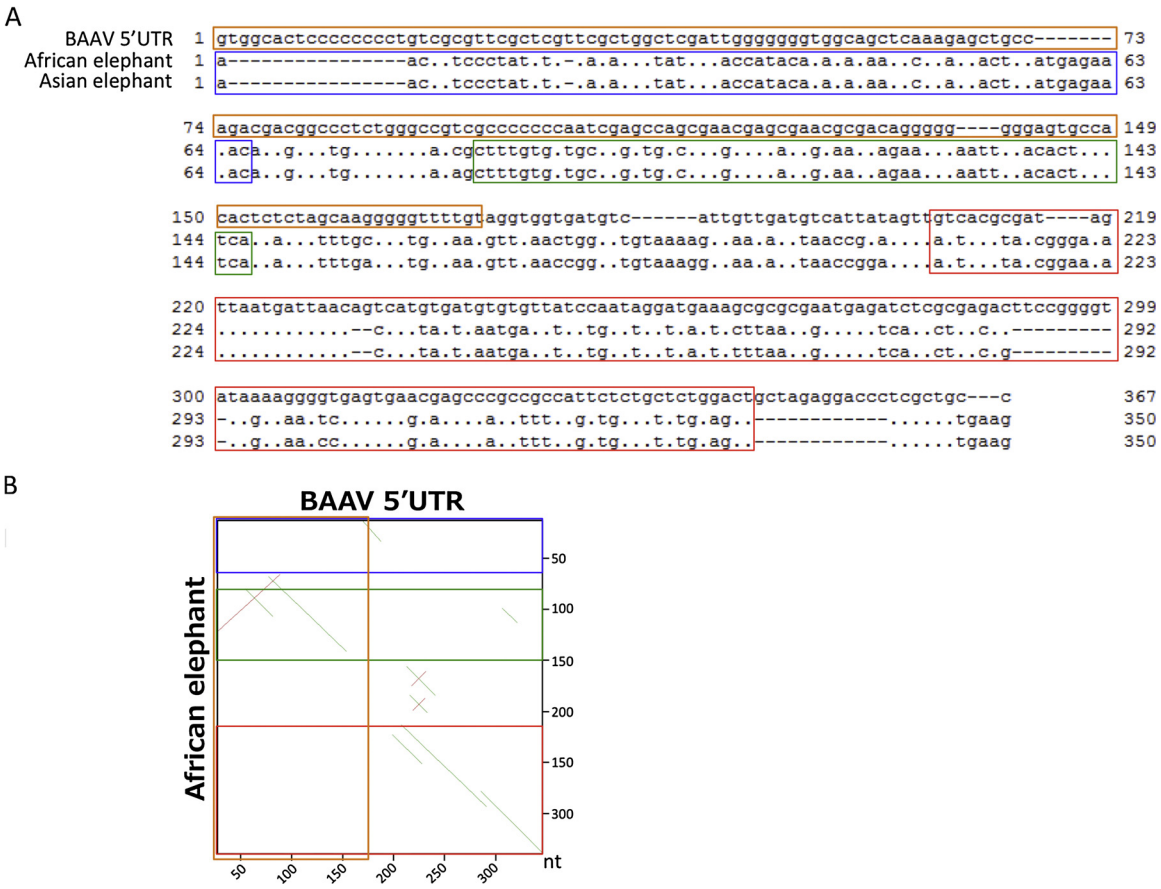

Fig. 5. Comparison of nucleotide sequences between the 5′ UTR of BAAV and 350 nt upstream from elephant enAAV-*rep*. (A) Multiple alignment of the nucleotide sequences of 5′ UTR of BAAV (AY388617) and 350 nt upstream from elephant enAAV-*rep*. (B) Dot-plot analysis of the 5′ UTR of BAAV and 350 nt upstream from African elephant enAAV-*rep*. Blue, green, orange, and red boxes indicate the positions of L1-2_LA, Gypsy147-I_DR, ITR, and the alignable region with the 5′ UTR of BAAV, respectively.

elephant enAAV-*rep*, suggesting that the enAAV-*rep* originated independently on the lineage of degu after divergence from the lineage of afrotherians (Fig. 3). These results suggest that the AAV Rep protein has been co-opted repeatedly in the mammalian liver. To understand the biological significance of these phenomena, it is important to clarify the functions of proteins encoded by enAAV-*rep* in elephants and degu.

Funding

This research was supported by a Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (No. 17H05824).

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

We thank Nobutaka Sato and Gen Bando at Asahiyama Zoo, who prepared elephant cells.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2018.04.015.

References

Arriagada, G., Gifford, R.J., 2014. Parvovirus-derived endogenous viral elements in two South American rodent genomes. J. Virol. 88, 12158–12162.
Aswad, A., Katzourakis, A., 2012. Paleovirology and virally derived immunity. Trends Ecol. Evol. 27, 627–636.
Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. Nat. Rev. Genet. 13, 283–296.
François, S., Filloux, D., Roumagnac, P., Bigot, D., Gayral, P., Martin, D.P., Froissart, R., Ogliastro, M., 2016. Discovery of parvovirus-related sequences in an unexpected broad range of animals. Sci. Rep. 6, 30880.
Fujino, K., Horie, M., Honda, T., Merriman, D.K., Tomonaga, K., 2014. Inhibition of Borna disease virus replication by an endogenous bornavirus-like element in the ground squirrel genome. Proc. Natl. Acad. Sci. U. S. A. 111, 13175–13180.
Gifford, R., Tristem, M., 2003. The evolution, distribution and diversity of endogenous retroviruses. Virus Genes 26, 291–315.
Hedges, S.B., Dudley, J., Kumar, S., 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972.
Henckaerts, E., Linden, R.M., 2010. Adeno-associated virus: a key to the human genome? Future Virol. 5, 555–574.
Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., Tomonaga, K., 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature 463, 84–87.
Horie, M., Kobayashi, Y., Suzuki, Y., Tomonaga, K., 2013. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. Philos. Trans. R. Soc. Lond. B Biol Sci. 368, 20120499.
Horie, M., Kobayashi, Y., Honda, T., Fujino, K., Akasaka, T., Kohl, C., Wibbelt, G., Mühldorfer, K., Kurth, A., Müller, M.A., Corman, V.M., Gillich, N., Suzuki, Y., Schwemmle, M., Tomonaga, K., 2016. An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. Sci. Rep. 6, 25873.
Kapoor, A., Simmonds, P., Lipkin, W.I., 2010. Discovery and characterization of mammalian endogenous parvoviruses. J. Virol. 84, 12628–12635.
Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6, e1001191.
Kobayashi, Y., Horie, M., Nakano, A., Murata, K., Itou, T., Suzuki, Y., 2016. Exaptation of Bornavirus-like nucleoprotein elements in Afrotherians. PLoS Pathog. 12, e1005785.
Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A.,

McLaughlin, S., Muzyczka, N., Rocchi, M., Berns, K.I., 1990. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. U. S. A. 87, 2211–2215.
Kotin, R.M., Menninger, J.C., Ward, D.C., Berns, K.I., 1991. Mapping and direct visualization of a region-specific viral DNA integration site on chromosome 19q13-qter. Genomics 10, 831–834.
Kohany, O., Gentles, A.J., Hankus, L., Jurka, J., 2006. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and censor. BMC Bioinf. 7, 474.
Kryukov, K., Ueda, M.T., Imanishi, T., Nakagawa, S., 2018. Systematic survey of non-retroviral virus-like elements in eukaryotic genomes. Virus Res (in press). https://www.sciencedirect.com/journal/virus-research/articles-in-press.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Myers, K.N., Barone, G., Ganesh, A., Staples, C.J., Howard, A.E., Beveridge, R.D., Maslen, S., Skehel, J.M., Collis, S.J., 2016. The bornavirus-derived human protein EBLN1 promotes efficient cell cycle transit, microtubule organisation and genome stability. Sci. Rep. 6, 35548.
Nishihara, H., Okada, N., 2008. Retroposons: genetic footprints on the evolutionary paths of life. Methods Mol. Biol. 422, 201–425.
Linden, R.M., Ward, P., Giraud, C., Winocour, E., Berns, K.I., 1996. Site-specific integration by adeno-associated virus. Proc. Natl. Acad. Sci. U. S. A. 93, 11288–11294.
Ling, C., Wang, Y., Lu, Y., Wang, L., Jayandharan, G.R., Aslanidi, G.V., Li, B., Cheng, B., Ma, W., Lentz, T., Ling, C., Xiao, X., Samulski, R.J., Muzyczka, N., Srivastava, A., 2015. The adeno-associated virus genome packaging puzzle. J. Mol. Genet. Med. 9, 175.
Rote, N.S., Chakrabarti, S., Stetzer, B.P., 2004. The role of human endogenous retroviruses in trophoblast differentiation and placental development. Placenta 25, 673–683.
Schmidt, M., Katano, H., Bossis, I., Chiorini, J.A., 2004. Cloning and characterization of a bovine adeno-associated virus. J. Virol. 78, 6509–6516.
Srivastava, A., Lusby, E.W., Berns, K.I., 1983. Nucleotide sequence and organization of the adeno-associated virus 2 genome. J. Virol. 45, 555–564.
Wang, X.S., Ponnazhagan, S., Srivastava, A., 1996. Rescue and replication of adeno-associated virus type 2 as well as vector DNA sequences from recombinant plasmids containing deletions in the viral inverted terminal repeats: selective encapsidation of viral genomes in progeny virions. J. Virol. 70, 1668–1677.