Virus Research 262 (2019) 2–9

Contents lists available at ScienceDirect

Virus Research

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

Review

Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses

Masayuki Horie[a,b,*], Keizo Tomonaga[b,c]

[a] Hakubi Center for Advanced Research, Kyoto University, Kyoto, Japan
[b] Department of Virus Research, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan
[c] Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto, Japan

ARTICLE INFO

Keywords:
Endogenous bornavirus-like elements
Endogenous viral elements
Paleovirology

ABSTRACT

Endogenous viral elements (EVEs) are virus-derived sequences embedded in eukaryotic genomes formed by germline integration of viral sequences. As many EVEs were integrated into eukaryotic genomes millions of years ago, EVEs are considered molecular fossils of viruses. EVEs can be valuable informational sources about ancient viruses, including their time scale, geographical distribution, genetic information, and hosts. Although integration of viral sequences is not required for replications of viruses other than retroviruses, many non-retroviral EVEs have been reported to exist in eukaryotes. Investigation of these EVEs has expanded our knowledge regarding virus–host interactions, as well as provided information on ancient viruses. Among them, EVEs derived from bornaviruses, non-retroviral RNA viruses, have been relatively well studied. Bornavirus-derived EVEs are widely distributed in animal genomes, including the human genome, and the history of bornaviruses can be dated back to more than 65 million years. Although there are several reports focusing on the biological significance of bornavirus-derived sequences in mammals, paleovirology of bornaviruses has not yet been well described and summarized. In this paper, we describe what can be learned about bornaviruses from endogenous bornavirus-like elements from the view of paleovirology using published results and our novel data.

1. Introduction

Bornaviruses (family *Bornaviridae*) are non-segmented negative strand RNA viruses that belong to the order *Mononegavirales* (Amarasinghe et al., 2017). Thus far, bornaviruses have been detected in snakes, birds, and several mammalian species, and are known to be the causative agents of fatal diseases such as Borna disease in horses and sheep and proventricular dilatation disease (PDD) or PDD-like diseases in birds (reviewed in (Amarasinghe et al., 2017; Kuhn et al., 2015)). Recently, several human cases of fatal encephalitis have been reported to be associated with infections by a bornavirus derived from squirrels (Hoffmann et al., 2015). In addition to the importance of bornaviruses as infectious agents, they are also known to be unusual viruses in that they establish persistent infections in the host-cell nucleus (Briese et al., 1992). Among known RNA viruses, the feature of nuclear persistence is observed in only two distinct lineages of mononegaviruses: bornaviruses and *Culex tritaeniorhynchus* rhabdovirus (CTRV) (Kuwata et al., 2011; Gillich et al., 2015) (Fig. 1). Thus, studies on these viruses may provide interesting insight into the evolution of RNA viruses, as well as the replication strategies used by RNA viruses.

Fossil records are powerful tools for studying evolution, because they provide information about prehistorical organisms, such as body size, body configuration, geographical distribution, and geological ages. Although viruses do not leave body fossils, retroviruses, which integrate their viral genomes into their host chromosomes during replication, are known to become endogenous following germline infection and integration (Weiss, 2006). The integrated viral genomes are called endogenous retroviruses (ERV). ERVs were discovered in the late 1960s to early 1970s, and as molecular fossils have provided important insights into past retroviruses and their hosts (Weiss, 2006).

Regarding bornaviruses, there are some descriptions about Borna disease-like illnesses in the older literature (reviewed in (Kuhn et al., 2015)), which imply the presence of bornaviruses more than a few hundred years ago. However, the presence of prehistorical RNA viruses, including bornaviruses, had been a mystery until relatively recently, due to the lack of evidence that RNA viruses existed. In 2010, we and others found that sequences homologous to bornaviruses are present in the genomes of animals, which are designated endogenous bornavirus-

Abbreviations: BoDV, Borna disease virus; EBL, endogenous bornavirus-like element; ERV, endogenous retrovirus; EVE, endogenous viral element; MTase, methyltransferase; MYA, million years ago; NYMV, Nyamanini virus; SbCNV-1, Soybean cyst nematode virus 1

* Corresponding author at: Department of Virus Research, Institute for Frontier Life and Medical Sciences, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo, Kyoto 606-8507, Japan.
E-mail address: horie.masayuki.3m@kyoto-u.ac.jp (M. Horie).

https://doi.org/10.1016/j.virusres.2018.04.006
Received 13 February 2018; Received in revised form 5 April 2018; Accepted 5 April 2018
Available online 06 April 2018
0168-1702/

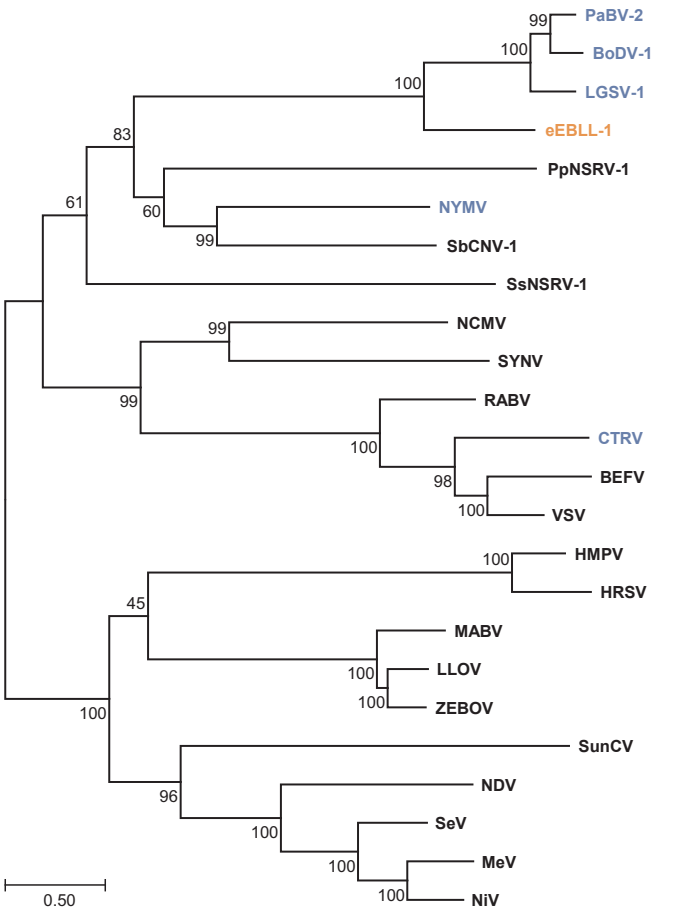

Fig. 1. A phylogenetic tree of RNA-dependent RNA polymerase genes of mononegaviruses. A phylogenetic tree was inferred by the Maximum Likelihood method based on the LG + G + I + F model and complete deletion option. The bootstrap values are shown for each interior branch. The scale bar indicates the number of amino acid substitutions per site. Virus replication in the host cell nucleus and an endogenous viral element are shown in blue and orange, respectively. PaBV-2: Parrot bornavirus 2, BoDV-1: Borna disease virus, LGSV-1: Loveridge's garter snake virus 1, eEBLL-1: *Eptesicus* endogenous bornavirus-like L 1, PpNSRV-1: *Pteromalus puparum* negative-strand RNA virus 1, NYMV: Nyamanini virus, SbCNV-1: Soybean cyst nematode virus 1, SsNSRV-1: *Sclerotinia sclerotiorum* negative-stranded RNA virus 1, NCMV: Northern cereal mosaic virus, SYNV: Sonchus yellow net virus, RABV: Rabies virus, CTRV: *Culex tritaeniorhynchus* rhabdovirus, BEFV: Bovine ephemeral fever virus, VSV: Vesicular stomatitis virus, HMPV: Human metapneumovirus, HRSV: Human respiratory syncytial virus, MABV: Marburg virus, LLOV: Lloviu cuevavirus, ZEBOV: Zaire ebolavirus, SunCV: Sunshine Coast virus, NDV: Newcastle disease virus, SeV: Sendai virus, MeV: Measles virus, NiV: Nipah virus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

like elements (EBLs) (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010). The discovery of EBLs drastically changed our knowledge about the paleovirology of RNA viruses, and it can provide important information regarding ancient bornaviruses, such as the hosts, geological ages, and genetic information. For example, from the analyses of EBLs in afrotherian animals, the history of bornaviruses can now be dated back to more than 65 million years ago (MYA) (Katzourakis and Gifford, 2010; Kobayashi et al., 2016). In addition, a novel concept that (partial) genomes of non-retroviral RNA viruses could be integrated into host chromosomes via retrotransposon enzymes was proposed (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010). Furthermore, some EBLs were suggested to perform biological roles in the hosts (Fujino et al., 2014; Horie, 2017; Sofuku et al., 2015; Myers et al., 2016; Parrish et al., 2015; Parrish and Tomonaga, 2016; Honda and Tomonaga, 2016; Horie et al., 2013). Thus, bornaviruses appear to have been involved in the evolution of their hosts.

Although there are several review papers describing the (possible) biological functions of EBLs in their host organisms (Horie, 2017; Honda and Tomonaga, 2016; Horie et al., 2013), the paleovirology of bornaviruses has not yet been well described. In this paper, we summarize and discuss the paleovirology of bornaviruses, and by using several examples, we show what can be learned from the molecular fossils of bornaviruses.

2. Integration mechanisms of bornavirus genes

Bornaviruses encodes six genes: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), large RNA-dependent RNA polymerase (L), and accessory protein (X) genes (Briese et al., 1994). Thus far, EBLN, EBLM, EBLG, and EBLL, which are derived from N, M, G, and L genes, respectively, have been reported to exist almost exclusively in vertebrate genomes (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010; Horie et al., 2013; Gilbert et al., 2014; Cui et al., 2014). Because bornaviruses encode neither reverse transcriptase nor integrase, non-bornaviral factors must mediate the reverse transcription and integration events. There are several candidates responsible for the reverse transcription in host cells (Horie et al., 2013): LTR-retrotransposons (including ERVs), non-LTR retrotransposons, telomerase, and co-infection of reverse transcribing viruses such as retroviruses. Among these, a non-LTR retrotransposon, LINE-1, is highly likely to be involved in some integration events of bornaviral genes for the following reasons. First, almost all EBLs are derived from single bornaviral genes (an exception is described in the section "*The features of ancient bornaviral genes*"), and some have downstream poly-A stretches (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010; Horie et al., 2013; Gilbert et al., 2014; Cui et al., 2014), suggesting that such EBLs are derived from viral mRNA. Second, some EBLs and poly-A stretches are flanked by short direct repeats, called target site duplications (TSDs) (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010). These observations are consistent with the footprint of LINE-1-mediated integration (Moran et al., 1996), and LINE-1 is indeed reported to be involved in the formation of processed pseudogenes by reverse-transcribing cellular mRNA (Maestre et al., 1995; Esnault et al., 2000; Babushok and Kazazian, 2007). Thus, some EBLs were probably integrated into the host chromosome via LINE-1, as reported for cellular processed pseudogenes. The above features are also confirmed by biological experiments using an extant bornavirus, Borna disease virus (BoDV). Although BoDV does not encode reverse-transcriptase, we demonstrated that BoDV mRNAs are integrated into the host cell chromosome, many of which are flanked by TSDs (Horie et al., 2010, 2013). Importantly, in anthropoids, the insertion dates of

EBLNs are consistent with the time when LINE-1 is considered to have been highly active (Ohshima et al., 2003; Horie and Tomonaga, 2011). Thus, LINE-1 was highly likely to have been involved in the endogenization of EBLs.

However, interestingly, an EBLN element in thirteen-lined ground squirrels (*Ictidomys tridecemlineatus* EBLN: itEBLN) seems to have been endogenized after the extinction of LINE-1, suggesting another reverse transcriptase was involved in its endogenization (Suzuki et al., 2014). Because itEBLN is flanked by several LTR retrotransposons, one of them might have been involved in the endogenization of itEBLN (Suzuki et al., 2014). It is reported that the genomic RNA of lymphocytic choriomeningitis virus (LCMV), a non-retroviral RNA virus, is reverse-transcribed and integrated into the host cell chromosome by the LTR retrotransposon IAP, or even by co-infection of the retrovirus, HIV (Geuking et al., 2009). Additionally, some viral sequences were suggested to be integrated into the host genomes through recombination with LTR retrotransposons in *Aedes aegypti* (Whitfield et al., 2017). Thus, it is plausible that some EBLs became endogenous by LTR retrotransposons, or through co-infection of exogenous retroviruses. It would be interesting to perform experiments using cells from a thirteen-lined ground squirrel as Geuking et al. did for LCMV (Geuking et al., 2009).

3. Integrated bornaviral genes and their distribution tendency

As previously reported, there is a characteristic tendency of integrated bornaviral genes. This includes the facts that many EBLNs and EBLLs are distributed in a wide range of animals, several EBLMs and EBLGs are found only in a few lineages of vertebrates, and no X or P-like EBLs have been reported (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010; Horie et al., 2013; Gilbert et al., 2014; Cui et al., 2014). Why is there such a tendency? There are several considerable factors that may explain it.

First, there are several biological factors such as the amount of viral transcripts and the preference of reverse-transcriptase. The amount of specific transcripts may affect the tendency as the more that viral transcripts are present in cells, the greater the chance they have to be integrated into the host chromosome. It was reported that there is a gradient of transcripts from the 3' to 5'–end of the mononegavirus genome (Whelan et al., 2004). Thus, transcripts of the N gene may be the most abundant in bornavirus-infected cells, which may therefore have more of a chance to be inserted into the host chromosomes compared with the other genes. Preference of reverse-transcriptase might also be involved in the tendency. Namely, N and L mRNA might be preferentially recognized by the responsible reverse transcriptase, although the precise mechanism is unclear.

Second, EBLNs and EBLLs may tend to confer selective advantages to the hosts following their endogenizations, whereas the other genes may not, or may even be deleterious for their hosts. This is plausible because the BoDV P protein is reported to be a pathogenic factor of BoDV. Transgenic mice expressing BoDV P show behavioral abnormalities (Honda et al., 2011; Kamitani et al., 2003). In addition, P protein can act as a decoy of a host phosphatase that is reported to be involved in aberrant neuronal function, modulation of epigenetic signaling, and impairment of neurogenesis (Prat et al., 2009; Bonnaud et al., 2015; Scordel et al., 2015). Furthermore, P protein interacts with HMGB-1, and inhibits a physiological function of HMGB-1 (Kamitani et al., 2001). In contrast, expression of BoDV N protein inhibits infections of BoDV without obvious neurological disease or behavioral abnormalities (Horie, 2017; Schneider et al., 2003; Rauer et al., 2004). Additionally, BoDV N is reported to be a major antigen of cytotoxic T-lymphocyte, which is likely to be responsible for the pathology of Borna disease (Planz and Stitz, 1999; Stitz et al., 2002). Thus, expression of an N-like protein could confer immune tolerance to bornaviral N protein, resulting in the suppression of the relevant inflammatory disease (Horie, 2017; Horie et al., 2013). Therefore, endogenization of the bornaviral N gene may confer resistant to exogenous bornavirus infections or to bornavirus-induced diseases, without deleterious effects.

Third, the methods used for detection of EBLs may contribute to the tendency. The length of query genes affects the detection sensitivity, because the threshold of tBLASTn screening to detect EVEs is usually determined by E value, and the length of query sequences is considered in the calculation of E value (NCBI BLAST FAQs; https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=FAQ). The L gene is the largest among genes encoded by bornaviruses, whereas X, P and M genes are relatively short. For example, for BoDV strain He/80/FR (accession number AJ311522), the open reading frames (ORFs) of L, X, P, and M consists of 5136, 264, 606, and 429 nucleotides, respectively. This results in EBLLs being more detectable by tBLASTn than the other genes.

Fourth, the evolutionary rate of each genes may be an important factor. Genes rapidly evolving in exogenous viruses are difficult to be detected by tBLASTn. A more rapidly changing gene may be missed by the search resulting in their prevalence being under represented.

It is probably that these factors described above together contribute toward the ultimate tendency observed. In the case of the X/P genes, they are shorter and seem to evolve faster than the other genes based on the fact that X/P genes are less conserved among exogenous bornaviruses (Kistler et al., 2008). Because we have demonstrated that X/P genes can be integrated into the host chromosome (Horie et al., 2013), the preference of reverse transcriptase cannot explain the lack of X/P-derived EBLs. Thus, EBLs derived from the X/P genes might be present, but not have been detected by tBLASTn. On the other hand, as described above, expression of P protein can be deleterious, and thus the endogenized X/P may not have been fixed in the population. Alternatively, ancient bornaviruses might not have encoded ancestral X/P genes. Clearly, further analyses are necessary to understand the ultimate reason for the tendency.

4. Putative hosts of ancient bornaviruses

EVEs are believed to be formed by germline integration of viral sequences. This means that the presence of EVEs reflects past viral infections, and may be useful in understanding host ranges of ancient viruses. Interestingly, EBLs are almost exclusively present in vertebrate genomes, and only a few EBLLs, which are very distantly related to bornaviral L genes, are reported to exist in invertebrates (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010; Horie et al., 2013; Gilbert et al., 2014; Cui et al., 2014; Horie and Tomonaga, 2011). On the other hand, many EVEs derived from other mononegaviruses are detected in lineages other than vertebrates (Katzourakis and Gifford, 2010; Theze et al., 2014; Kondo et al., 2013; Fort et al., 2012; Geisler and Jarvis, 2016). These observations suggest that bornaviruses may have primarily infected vertebrate animals. Indeed, to date, exogenous bornaviruses have been detected only in vertebrates, such as mammals, birds, and snakes (Amarasinghe et al., 2017). Although we cannot exclude the possibility that bornaviral sequences are for some reason more difficult to be endogenized in other eukaryotic lineages, the hosts of bornaviruses might have been limited to vertebrate animals for a long time.

Although the distribution of EBLs are mostly limited to vertebrate animals, many of the lineages that have EBLs are not reported to be hosts for exogenous bornaviruses, such as afrotherian animals and bats. This suggests that there may be undiscovered exogenous bornaviruses in these lineages, or in closely related animal lineages. As animals possessing EBLs are believed to be less susceptible to bornavirus infection and bornaviral disease (Belyi et al., 2010), it would be interesting to investigate animals that do not have EBLs, but that are genetically similar to those animals with EBLs. In such cases, sequence information regarding the EBLs would be useful in designing primers for detecting exogenous bornaviruses using methods previously described (Horie, 2017).

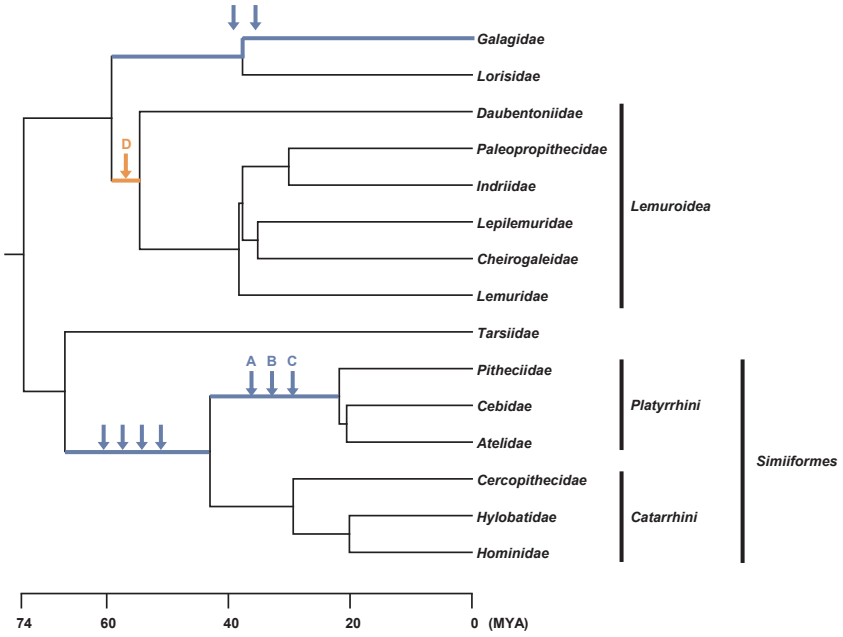

Fig. 2. Insertion dates of EBLs in primates. A schematic phylogenetic tree of primates and integration of EBLs are shown. The tree topology and time scale were adopted from TimeTree (Hedges et al., 2006). The blue and orange arrows indicate the integration events of EBLNs and an EBLN-EBLM, respectively. The characters on the arrows (A–D) are corresponding to the gene orthology analyses shown in Supplementary Fig. 1A–D. MYA, million years ago. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

5. The ages of bornaviruses

Insertion dates of EVEs can be estimated by determining gene orthologies (Aiewsakun and Katzourakis, 2015). Insertion dates of some EBLs have been reported. For example, some anthropoid EBLNs are reported to have integrated into the host genomes before the divergence of *Platyrrhini* and *Catarrhini*, which is estimated to be 43 MYA (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010). Similarly, an afrotherian EBLN and a bat EBLL are reported to have been integrated into the host genomes at least 65 and 11.8 MYA, respectively (Katzourakis and Gifford, 2010; Kobayashi et al., 2016; Horie et al., 2016).

In addition to the above EBLs, we have analyzed the integration date of primate EBLs as an example to study paleovirology of bornaviruses. The expansion of publicly available genomic sequence data has made it possible to detect more EBLs than was previously possible. This in turn makes it possible to determine the ages of additional EBLs. Thus, we have again screened EBLs by tBLASTn (the method and data regarding the detected EBLs are available in the Supplementary information). Among the detected EBLs, we have comprehensively determined the ages of EBLs in primates (Fig. 2 and Supplementary Fig. 1). Referring divergent times from Time Tree (Hedges et al., 2006), our analyses showed that four EBLNs became endogenous before the divergence of anthropoids (43 MYA) as previously described (Belyi et al., 2010; Horie et al., 2010; Katzourakis and Gifford, 2010); three EBLNs were inserted before the divergence of platyrrhines (21.9 MYA); and an EBLN and EBLM were formed before the divergence of the *Lemuroidea* (55 MYA). The ages of two EBLNs in *Otolemur garnettii* were difficult to determine due to the lack of genomic data for species genetically similar to *O. garnettii*. In this case, genomic PCR amplification and sequencing are necessary for determining the ages; the approach we and others previously used for analysis an EBLL in bats and EBLNs in snakes (Gilbert et al., 2014; Horie et al., 2016). Taken together, these findings suggest that bornaviruses have infected a wide range of ancestral primates.

It was proposed that some EBLs may function and/or have functioned as anti-bornaviral genes because of the tendency that animals having EBLs are less susceptible to bornavirus infections and bornaviral diseases (Belyi et al., 2010). Indeed, we showed that exogenous expression of itEBLN inhibits BoDV replication (Fujino et al., 2014). Additionally, antisense piRNAs are expressed from some EBL loci, which might act or have acted as antiviral small RNAs (Parrish et al., 2015; Honda and Tomonaga, 2016). Also, several possible anti-bornaviral mechanisms of EBLs have been proposed as described above (the Section 3 *Integrated bornaviral genes and their distribution tendency*), and especially EBLNs might act as anti-bornaviral genes for a long period (Horie, 2017). However, interestingly, we found that three EBLNs in platyrrhine animals became endogenous after the earlier fixation of four EBLNs. Gilbert et al. reported a similar situation in snake EBLNs (Gilbert et al., 2014). Thus, the previously integrated EBLNs did not inhibit subsequent bornaviral infections. It could have been that the previously integrated EBLNs were not expressed at the time of the subsequent bornaviral infections or were not be able to inhibit the later infections of bornaviruses even if they were expressed. This means that the co-evolution between bornaviruses and their hosts is more complex than previously thought. Comprehensive analysis of EBLs would provide more insight into the relationship between bornaviruses and their hosts.

6. Features of ancient bornaviral genes

EVEs can provide genetic information about ancient viruses. In this section, we describe two examples of EBLs that would be useful in the understanding of genetic information about ancient bornaviruses.

An EBLL in bats of the genus *Eptesicus*, named eEBLL-1, retains a large ORF consisting of 1718 codons, which is almost identical in length to the L genes of modern bornaviruses, approximately 1710 codons (Horie et al., 2016). Because orthologous genes were detected from three species of bats of the genus *Eptesicus* (*E. fuscus*, *E. nilssonii*, and *E. serotinus*), eEBLL-1 was most likely integrated into the genome prior to the divergence of the three species, which is estimated to be 11.8 MYA (Horie et al., 2016). A TSD and poly-A stretch are observed near the eEBLL-1, suggesting that the ORF of eEBLL-1 reflects the full-length ORF of an ancient bornavirus (Horie et al., 2016), and thus eEBLL-1 would provide precious information regarding an ancient bornaviral L gene as described below.

As reported for modern bornavirus L genes, the ORF of eEBLL-1 is shorter than the L genes of other mononegaviruses (Horie et al., 2016), and lacks block VI, which is one of the conserved regions among mononegaviruses (Briese et al., 1994; Poch et al., 1990; Ogino and Banerjee, 2011) (Fig. 3). Block VI is part of methyl transferase (MTase) domain that is responsible for the methylation of mRNA cap, which is

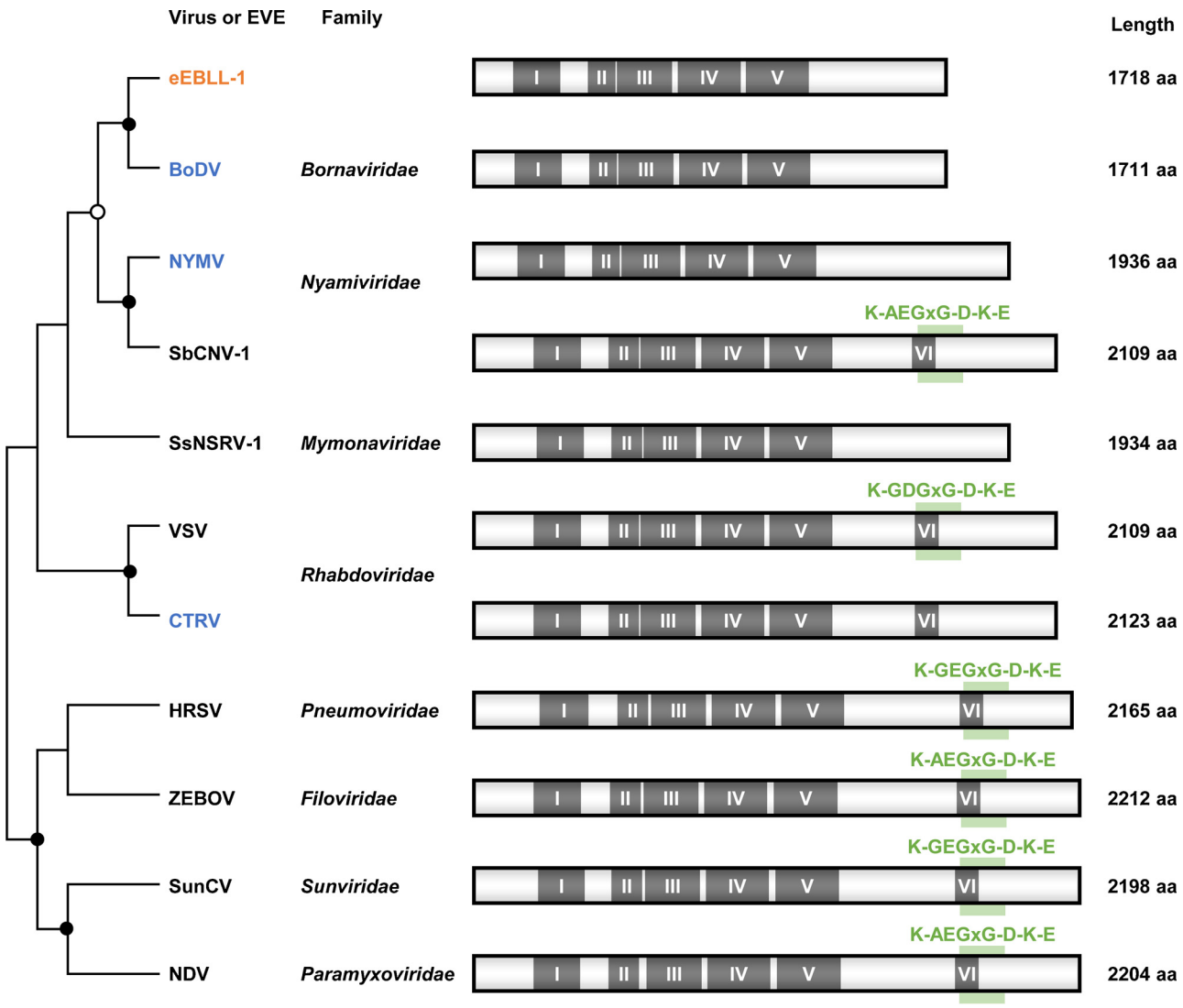

Fig. 3. Structures of L proteins of mononegaviruses. A schematic structure of L or EBLL proteins and the phylogenetic tree are shown. The tree topology was adopted from Fig. 1. The black boxes with I–VI show the highly conserved boxes among the L proteins (Poch et al., 1990). The MTase domain and each motif sequence are shown in green. The black and white circles indicate the bootstrap values more than 95 and more than 80, respectively. Blue and orange show the viruses replicating in the nucleus and EBLL, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

critical for translation, splicing, polyadenylation, mRNA transport, and escape from detection by the innate immune system (reviewed in (Decroly et al., 2011)). The capping reaction of cellular mRNAs including the methylation is performed in the cell nucleus. A large part of mononegaviruses are transcribed in the cytoplasm, and thus the viruses possess their own MTase, which allows for the efficient expression of the viral genes and the escape from the cellular immune system (Decroly et al., 2011). On the other hand, bornaviruses transcribe their mRNAs in the cell nucleus (Briese et al., 1992). Therefore, bornaviruses may exploit the host MTase in the nucleus, and thus bornaviral L protein does not have, nor need the MTase domain (Ogino, 2014). Interestingly, L proteins of Nyamanini virus (NYMV) and CTRV, which were reported to replicate and be transcribed in the nucleus (Briese et al., 1992), also lack the MTase domain (Ogino, 2014). Although the authentic root of mononegaviruses is unclear, a midpoint root tree suggests that these nuclear-replicating viruses have lost the MTase domain during evolution (Fig. 3). Thus, the lack of an MTase domain may be an indicator of nuclear replication of mononegaviruses. Intriguingly, L protein of soybean cyst nematode virus 1 (SbCNV-1), a virus belonging to the same family as NYMV (the family *Nyamiviridae*) (Bekal et al., 2011), possesses the MTase domain (Fig. 3). Thus, the three lineages of nuclear-replicating viruses may have lost the MTase domain convergently. However, the time when the viruses invaded the nucleus is unknown. Probably reflecting a full-length L gene of ancient bornavirus, eEBLL-1 also lacks block VI and the MTase domain. Therefore, the transcription and replication of an ancient bornavirus, which infected a common ancestor of bats of the genus *Eptesicus*, may have been performed in the nucleus at least 11.8 MYA. Additionally, the truncation of the C-terminal region of L protein occurred at least 11.8 MYA. As described here, much can be learned regarding the evolution of viruses using sequences of EBLs.

We can also learn about the genome organization of ancient bornaviruses. During the gene orthology analyses in this study, we found that an EBLN and EBLM were closely located in the genome of the gray mouse lemur *Microcebus murinus* (Fig. 4). The orthologous EBLs are present in members of the suborder *Lemuroidea* including *Eulemur macaco* (black lemur), *E. flavifrons* (blue-eyed black lemur), *Daubentonia madagascariensis* (aye-aye), and *Propithecus coquereli* (Coquerel's sifaka) (Supplementary Fig. 1). Thus, the EBLs likely became endogenous at some point before the divergence of the *Lemuroidea* animals, which is

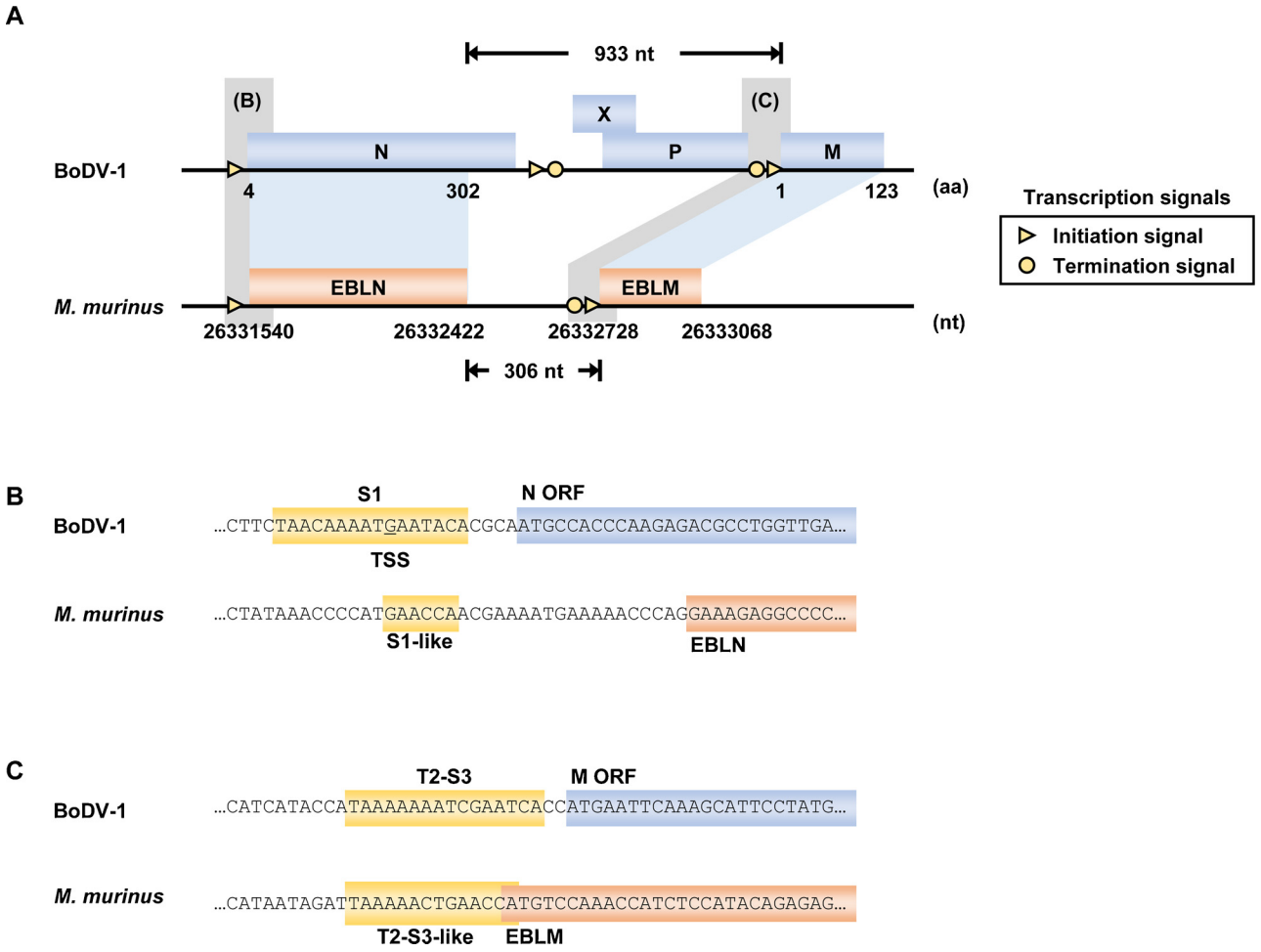

Fig. 4. Nucleotide sequence alignment between BoDV genome and the EBLN-EBLM region in *Microcebus mirinus*. **(A)** A schematic diagram of the BoDV genome and the EBLN-EBLM region in *M. murinus*. Transcription signals are indicated with yellow triangles and circles. The numbers below the BoDV-1 genome shows the amino acid positions within each gene. The numbers below the *M. murinus* genome is the nucleotide positions in the contig NC_033661. The regions shown in (B) and (C) are indicated in gray. The BLAST hit regions are shown in light blue. nt, nucleotides; aa, amino acids. **(B)** and **(C)** Nucleotide sequence alignments of BoDV and *M. murinus* genomes. The alignment regions are indicated in **(A)**. Yellow boxes show transcription initiation (S1 and S3) or termination (T2) signals and S1 or T2-S3-like sequence. Blue and orange boxes indicate the open reading frame of BoDV N or M and EBLN or EBLM regions, respectively. The transcription start site (TSS) in S1 is indicated with the underline. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

estimated to be 55 MYA (Hedges et al., 2006). This element is interesting because the sequences of these EBLs may reflect the genome structure of an ancient bornavirus that infected the common ancestor of the modern-day *Lemuroidea* animals.

In the genome of *M. murinus*, the EBLN and EBLM correspond to amino acid residues 4-302 and 1-123 of BoDV N and M, respectively (Fig. 4A). A bornaviral transcription initiation signal-like sequence is present immediately upstream of the EBLN (Fig. 4A and B), and transcription termination and initiation signal-like sequences are also observed immediately upstream of the EBLM (Fig. 4A and C). The length between the EBLN and EBLM is 306 nucleotides (nt), which is much shorter than that of the corresponding region in the BoDV genome (933 nt). Although there is an A-rich region downstream of the EBLN, this A-rich region may not be a poly-A tail, as a similar A-rich region is also present in the counterpart of the BoDV genome (Supplementary Fig. 2). Moreover, the orthologues have the same genomic structure in this region, except for *P. coquereli*, which contains long undetermined nucleotides (N) between the EBLN and EBLM region, suggesting that the structure of the common ancestral sequence is similar to that of *M. murinus*. Taken together, these observations suggest that the genome organization of the ancient bornavirus is significantly different from that of modern exogenous bornaviruses. There are several possibilities: first, in the ancient bornavirus, the order of genes is different from that of modern viruses, second, the X/P genes are absent, or third, the X/P genes are very short. Interestingly, there is only one transcription initiation-termination signal-like sequences between the EBLN and EBLM (Fig. 4A and C), supporting the first two possibilities. Further analyses are required to assess these possibilities. It would be of interest to comprehensively analyze the EBLs and other non-retroviral EVEs to understand the evolution of viral genomes.

7. Conclusion and perspectives

Thanks to the recent progress in sequencing techniques, the amount and availability of public genome data are rapidly expanding. However, there are only a few studies that have deeply analyzed non-retroviral EVEs from the view of paleovirology. Detailed phylogenetic analyses of EVEs could provide information about transmission routes and geographical distributions of ancient viruses (Aiewsakun and Katzourakis, 2015). Moreover, application of molecular biological approaches should prove useful toward understanding ancient viruses. A recent report showed reconstruction of a protein of ancient adeno-associated virus (Smith et al., 2016). Such an approach would be beneficial in the assessment of the functionality of ancient viral proteins. The combination of these methods will lead to a deeper understanding of RNA viruses.

Declarations of interest

None.

Acknowledgements

We thank to Dr. Keiko Takemoto (Institute for Frontier Life and Medical Sciences, Kyoto University) for technical assistances. We are grateful to Bea Clarise B.Garcia for useful discussion. This work is supported by the Hakubi project (MH) and a Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports, and Technology (MEXT) of Japan [No. 16H06429 (TK), 16H06430 (TK), 16K21723 (MH and TK), and 17H05821 (MH)], JSPS KAKENHI grant numbers JP26253027 (KT) and 26850208 (MH), JSPS Core-to-Core Program, Advanced Research Networks (KT), and AMED grant number (JP17fm0208014 and JP17am0301015h0003) (KT).

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.2018.04.006.

References

Aiewsakun, P., Katzourakis, A., 2015. Endogenous viruses: connecting recent and ancient viral evolution. Virology 479–480, 26–37.
Amarasinghe, G.K., Bao, Y., Basler, C.F., Bavari, S., Beer, M., Bejerman, N., Blasdell, K.R., Bochnowski, A., Briese, T., Bukreyev, A., et al., 2017. Taxonomy of the order Mononegavirales: update 2017. Arch. Virol. 162, 2493–2504.
Babushok, D.V., Kazazian Jr., H.H., 2007. Progress in understanding the biology of the human mutagen LINE-1. Hum. Mutat. 28, 527–539.
Bekal, S., Domier, L.L., Niblack, T.L., Lambert, K.N., 2011. Discovery and initial analysis of novel viral genomes in the soybean cyst nematode. J. Gen. Virol. 92, 1870–1879.
Belyi, V.A., Levine, A.J., Skalka, A.M., 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS Pathog. 6, e1001030.
Bonnaud, E.M., Szelechowski, M., Betourne, A., Foret, C., Thouard, A., Gonzalez-Dunia, D., Malnou, C.E., 2015. Borna disease virus phosphoprotein modulates epigenetic signaling in neurons to control viral replication. J. Virol. 89, 5996–6008.
Briese, T., de la Torre, J.C., Lewis, A., Ludwig, H., Lipkin, W.I., 1992. Borna disease virus, a negative-strand RNA virus, transcribes in the nucleus of infected cells. Proc. Natl. Acad. Sci. U. S. A. 89, 11486–11489.
Briese, T., Schneemann, A., Lewis, A.J., Park, Y.S., Kim, S., Ludwig, H., Lipkin, W.I., 1994. Genomic organization of Borna disease virus. Proc. Natl. Acad. Sci. U. S. A. 91, 4362–4366.
Cui, J., Zhao, W., Huang, Z., Jarvis, E.D., Gilbert, M.T., Walker, P.J., Holmes, E.C., Zhang, G., 2014. Low frequency of paleoviral infiltration across the avian phylogeny. Genome Biol. 15, 539.
Decroly, E., Ferron, F., Lescar, J., Canard, B., 2011. Conventional and unconventional mechanisms for capping viral mRNA. Nat. Rev. Microbiol. 10, 51–65.
Esnault, C., Maestre, J., Heidmann, T., 2000. Human LINE retrotransposons generate processed pseudogenes. Nat. Genet. 24, 363–367.
Fort, P., Albertini, A., Van-Hua, A., Berthomieu, A., Roche, S., Delsuc, F., Pasteur, N., Capy, P., Gaudin, Y., Weill, M., 2012. Fossil rhabdoviral sequences integrated into arthropod genomes: ontogeny, evolution, and potential functionality. Mol. Biol. Evol. 29, 381–390.
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.
Geisler, C., Jarvis, D.L., 2016. Rhabdovirus-like endogenous viral elements in the genome of Spodoptera frugiperda insect cells are actively transcribed: implications for adventitious virus detection. Biologicals 44, 219–225.
Geuking, M.B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., Zinkernagel, R.M., Hangartner, L., 2009. Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. Science 323, 393–396.
Gilbert, C., Meik, J.M., Dashevsky, D., Card, D.C., Castoe, T.A., Schaack, S., 2014. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. Proc. Biol. Sci. 281, 1122.
Gillich, N., Kuwata, R., Isawa, H., Horie, M., 2015. Persistent natural infection of a Culex tritaeniorhynchus cell line with a novel Culex tritaeniorhynchus rhabdovirus strain. Microbiol. Immunol. 59, 562–566.
Hedges, S.B., Dudley, J., Kumar, S., 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22, 2971–2972.
Hoffmann, B., Tappe, D., Hoper, D., Herden, C., Boldt, A., Mawrin, C., Niederstrasser, O., Muller, T., Jenckel, M., van der Grinten, E., et al., 2015. A variegated squirrel bornavirus associated with fatal human encephalitis. N. Engl. J. Med. 373, 154–162.
Honda, T., Tomonaga, K., 2016. Endogenous non-retroviral RNA virus elements evidence a novel type of antiviral immunity. Mob. Genet. Elem. 6, e1165785.
Honda, T., Fujino, K., Okuzaki, D., Ohtaki, N., Matsumoto, Y., Horie, M., Daito, T., Itoh, M., Tomonaga, K., 2011. Upregulation of insulin-like growth factor binding protein 3 in astrocytes of transgenic mice that express Borna disease virus phosphoprotein. J. Virol. 85, 4567–4571.
Horie, M., 2017. The biological significance of bornavirus-derived genes in mammals. Curr. Opin. Virol. 25, 1–6.
Horie, M., Tomonaga, K., 2011. Non-retroviral fossils in vertebrate genomes. Viruses 3, 1836–1848.
Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J.M., et al., 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., Muhldorfer, K., Kurth, A., Muller, M.A., et al., 2016. An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. Sci. Rep. 6, 25873.
Kamitani, W., Ono, E., Yoshino, S., Kobayashi, T., Taharaguchi, S., Lee, B.J., Yamashita, M., Kobayashi, T., Okamoto, M., Taniyama, H., et al., 2003. Glial expression of Borna disease virus phosphoprotein induces behavioral and neurological abnormalities in transgenic mice. Proc. Natl. Acad. Sci. U. S. A. 100, 8969–8974.
Kamitani, W., Shoya, Y., Kobayashi, T., Watanabe, M., Lee, B.J., Zhang, G., Tomonaga, K., Ikuta, K., 2001. Borna disease virus phosphoprotein binds a neurite outgrowth factor, amphoterin/HMG-1. J. Virol. 75, 8742–8751.
Katzourakis, A., Gifford, R.J., 2010. Endogenous viral elements in animal genomes. PLoS Genet. 6, e1001191.
Kistler, A.L., Gancz, A., Clubb, S., Skewes-Cox, P., Fischer, K., Sorber, K., Chiu, C.Y., Lublin, A., Mechani, S., Farnoushi, Y., et al., 2008. Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. Virol. J. 5, 88.
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.
Kondo, H., Chiba, S., Toyoda, K., Suzuki, N., 2013. Evidence for negative-strand RNA virus infection in fungi. Virology 435, 201–209.
Kuhn, J.H., Durrwald, R., Bao, Y., Briese, T., Carbone, K., Clawson, A.N., deRisi, J.L., Garten, W., Jahrling, P.B., Kolodziejek, J., et al., 2015. Taxonomic reorganization of the family Bornaviridae. Arch. Virol. 160, 621–632.
Kuwata, R., Isawa, H., Hoshino, K., Tsuda, Y., Yanase, T., Sasaki, T., Kobayashi, M., Sawabe, K., 2011. RNA splicing in a new rhabdovirus from Culex mosquitoes. J. Virol. 85, 6185–6196.
Maestre, J., Tchenio, T., Dhellin, O., Heidmann, T., 1995. mRNA retroposition in human cells: processed pseudogene formation. EMBO J. 14, 6333–6338.
Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., Kazazian Jr., H.H., 1996. High frequency retrotransposition in cultured mammalian cells. Cell 87, 917–927.
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.
Ogino, T., 2014. The multifunctional RNA polymerase L protein of non-segmented negative strand RNA viruses catalyzes unique mRNA capping. Uirusu 64, 165–178.
Ogino, T., Banerjee, A.K., 2011. An unconventional pathway of mRNA cap formation by vesiculoviruses. Virus Res. 162, 100–109.
Ohshima, K., Hattori, M., Yada, T., Gojobori, T., Sakaki, Y., Okada, N., 2003. Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates. Genome Biol. 4, R74.
Parrish, N.F., Tomonaga, K., 2016. Endogenized viral sequences in mammals. Curr. Opin. Microbiol. 31, 176–183.
Parrish, N.F., Fujino, K., Shiromoto, Y., Iwasaki, Y.W., Ha, H., Xing, J., Makino, A., Kuramochi-Miyagawa, S., Nakano, T., Siomi, H., et al., 2015. piRNAs derived from ancient viral processed pseudogenes as transgenerational sequence-specific immune memory in mammals. RNA 21, 1691–1703.
Planz, O., Stitz, L., 1999. Borna disease virus nucleoprotein (p40) is a major target for CD8(+)-T-cell-mediated immune response. J. Virol. 73, 1715–1718.
Poch, O., Blumberg, B.M., Bougueleret, L., Tordo, N., 1990. Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. 71 (Pt. 5), 1153–1162.
Prat, C.M., Schmid, S., Farrugia, F., Cenac, N., Le Masson, G., Schwemmle, M., Gonzalez-Dunia, D., 2009. Mutation of the protein kinase C site in Borna disease virus phosphoprotein abrogates viral interference with neuronal signaling and restores normal synaptic activity. PLoS Pathog. 5, e1000425.
Rauer, M., Gotz, J., Schuppli, D., Staeheli, P., Hausmann, J., 2004. Transgenic mice expressing the nucleoprotein of Borna disease virus in either neurons or astrocytes: decreased susceptibility to homotypic infection and disease. J. Virol. 78, 3621–3632.
Schneider, U., Naegele, M., Staeheli, P., Schwemmle, M., 2003. Active Borna disease virus polymerase complex requires a distinct nucleoprotein-to-phosphoprotein ratio but no viral X protein. J. Virol. 77, 11781–11789.
Scordel, C., Huttin, A., Cochet-Bernoin, M., Szelechowski, M., Poulet, A., Richardson, J., Benchoua, A., Gonzalez-Dunia, D., Eloit, M., Coulpier, M., 2015. Borna disease virus phosphoprotein impairs the developmental program controlling neurogenesis and reduces human GABAergic neurogenesis. PLoS Pathog. 11, e1004859.
Smith, R.H., Hallwirth, C.V., Westerman, M., Hetherington, N.A., Tseng, Y.S., Cecchini, S., Virag, T., Ziegler, M.L., Rogozin, I.B., Koonin, E.V., et al., 2016. Germline viral

"fossils" guide in silico reconstruction of a mid-Cenozoic era marsupial adeno-associated virus. Sci. Rep. 6, 28965.
Sofuku, K., Parrish, N.F., Honda, T., Tomonaga, K., 2015. Transcription profiling demonstrates epigenetic control of non-retroviral RNA virus-derived elements in the human genome. Cell Rep. 12, 1548–1554.
Stitz, L., Bilzer, T., Planz, O., 2002. The immunopathogenesis of Borna disease virus infection. Front. Biosci. 7, d541–555.
Suzuki, Y., Kobayashi, Y., Horie, M., Tomonaga, K., 2014. Origin of an endogenous bornavirus-like nucleoprotein element in thirteen-lined ground squirrels. Genes Genet. Syst. 89, 143–148.
Theze, J., Leclercq, S., Moumen, B., Cordaux, R., Gilbert, C., 2014. Remarkable diversity of endogenous viruses in a crustacean genome. Genome Biol. Evol. 6, 2129–2140.
Weiss, R.A., 2006. The discovery of endogenous retroviruses. Retrovirology 3, 67.
Whelan, S.P., Barr, J.N., Wertz, G.W., 2004. Transcription and replication of non-segmented negative-strand RNA viruses. Curr. Top. Microbiol. Immunol. 283, 61–119.
Whitfield, Z.J., Dolan, P.T., Kunitomi, M., Tassetto, M., Seetin, M.G., Oh, S., Heiner, C., Paxinos, E., Andino, R., 2017. The diversity, structure, and function of heritable adaptive immunity sequences in the Aedes aegypti genome. Curr. Biol. 27, 3511–3519 e3517.