Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes

Masayuki Horie1,2, Yuki Kobayashi3, Yoshiyuki Suzuki4 and Keizo Tomonaga2

1Department of Virology, Institute for Medical Microbiology and Hygiene, University of Freiburg, 79104 Freiburg, Germany
2Department of Viral Oncology, Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan
3Nihon University Veterinary Research Center, Fujisawa, Kanagawa 252-8510, Japan
4Graduate School of Natural Sciences, Nagoya City University, Nagoya, Aichi 467-8501, Japan

Bornaviruses are the only animal RNA viruses that establish a persistent infection in their host cell nucleus. Studies of bornaviruses have provided unique information about viral replication strategies and virus–host interactions. Although bornaviruses do not integrate into the host genome during their replication cycle, we and others have recently reported that there are DNA sequences derived from the mRNAs of ancient bornaviruses in the genomes of vertebrates, including humans, and these have been designated endogenous borna-like (EBL) elements. Therefore, bornaviruses have been interacting with their hosts as driving forces in the evolution of host genomes in a previously unexpected way. Studies of EBL elements have provided new models for virology, evolutionary biology and general cell biology. In this review, we summarize the data on EBL elements including what we have newly identified in eukaryotes genomes, and discuss the biological significance of EBL elements, with a focus on EBL nucleoprotein elements in mammalian genomes. Surprisingly, EBL elements were detected in the genomes of invertebrates, suggesting that the host range of bornaviruses may be much wider than previously thought. We also review our new data on non-retroviral integration of Borna disease virus.

1. Exogenous bornaviruses

Bornaviruses are enveloped, non-segmented, negative-strand RNA viruses in the order Mononegavirales [1]. Borna disease virus (BDV) is a mammalian bornavirus and the type species of the genus Bornavirus in the family Bornaviridae. BDV was originally identified as the causative agent of Borna disease, which is characterized by non-purulent meningoencephalomyelitis of horse and sheep. Borna disease was named after the town Borna, Germany, where there was a significant epidemic of the disease at the end of the nineteenth century. The history of Borna disease traces back to the seventeenth century when Borna disease was described as ‘disease of the head’ [2].

Bornaviruses preferentially infect nervous system cells in vivo without obvious cell death. The most remarkable feature of bornavirus pathology is persistent infection in the host cell nucleus. Among the animal RNA viruses, viruses in only three taxa are known to replicate in the nucleus: the families Bornaviridae and Orthomyxoviridae and the genus Nyavirus, although the Nyavirus Midway virus has not yet been shown to replicate in the cell nucleus [1,3,4]. Because bornaviruses establish non-cytolytic persistent infections and viruses in the other two taxa cause lytic infections, bornaviruses are the only RNA viruses that have been reported to produce persistent infections in the host cell nucleus. We recently showed that BDV establishes an intranuclear persistent infection by attaching to host chromatin [5]. Thus, bornaviruses present interesting viral replication strategies and virus–host interactions.

The bornavirus genome is an 8.9 kb minus-strand RNA, which encodes six genes: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein...
RNA-dependent RNA polymerase (L) and accessory protein (X; figure 1). N encapsidates the viral RNA to form the viral nucleocapsid. P is a cofactor of viral polymerase L and is also a phosphorylation decoy involved in BDV pathogenicity [6–13]. Viral gene products N, P and L are the minimal components of viral ribonucleoprotein (vRNP) [14]. M and G are structural proteins. M is thought to line the inside of virions. Interestingly, M also associates with vRNP in the host cell nucleus, suggesting that M is involved in viral replication or transport of viral components [15]. G is the viral envelope glycoprotein and is involved in BDV entry, involving virion attachment to an unknown receptor and fusion of the viral envelope and cell membrane to release the vRNP into the cell cytoplasm in association with host factors [16–21]. X is a multifunctional, non-structural protein that is essential for the viral replication cycle [22], and is known to be a regulator of viral polymerase activity and an inhibitor of apoptosis in the central nervous system [23–25].

BDV has been reported to infect a wide range of mammalian species and several avian species [26–28]. BDV was suggested to be involved in a human psychiatric disorder in 1985 [29,30], but several more recent reports have challenged this idea [31]. Although sporadic BDV infections are still observed in several animals, the natural reservoir of BDV has not yet been identified. The bicoloured white-toothed shrew (*Crocidura leucodon*) is a putative natural host of BDV in central Europe [32,33]. However, in view of the range of its habitat, this does not explain the natural reservoir of BDV in other regions and suggests there must be other reservoirs [34].

Interestingly, the nucleotide sequences of BDV isolates show extremely high conservation among BDV strains, independent of host species, and region and year of isolation, except for strain no/98 [35]. Therefore, until recently, BDV was thought to be the only member of the family Bornaviridae.

However, avian bornavirus (ABV), a new member of the family *Bornaviridae*, was isolated from psittacine birds with proventricular dilatation disease (PDD) in 2008 [36,37]. PDD is a fatal neurological disorder characterized by lymphoplasmacytic inflammation of the central and peripheral nervous system [38]. Because Koch’s postulates were experimentally fulfilled, ABV has been considered the aetiological agent of PDD [39,40]. However, cases of asymptomatic infection have been reported occasionally, indicating that ABV infection does not always cause PDD and raising the importance of epidemiological studies of ABV [41–43]. In addition to psittacine birds, ABV has been isolated from non-psittacine birds, i.e. canary (*Serinus canaria*), Canada goose (*Branta canadensis*), trumpeter swan (*Cygnus buccinator*) and mute swan (*Cygnus olor*) [44–47]. Nine ABV genotypes have been identified thus far, with seven in psittacine species, one in a canary and one in the Canada goose and swans [37,48,49]. In addition, we found ABV-like sequences in the expressed sequence tags (ESTs) database of Bengalese finch (*Lonchura striata domestica*) by BLAST analysis. Because these ABV-like sequences almost completely cover all ABV mRNAs, they are probably ABV transcripts. In addition, phylogenetic analysis has shown that the sequences are genetically distinct from known ABV genotypes, suggesting that they are a novel ABV genotype [50].
In addition to BDV and ABV, bornavirus N- and X/P-like sequences were detected in a cDNA library derived from the venom gland, but not the genomic DNA, of a Gaboon viper (Bitis gabonica), suggesting that there may be a reptile bornavirus (RBV) although such a virus has not yet been isolated [51]. BDV, ABV and RBV enable research involving comparison of bornavirus nucleotide and amino acid sequences, which cannot be performed with only the highly conserved BDV nucleotide sequences. Such studies provide us with interesting insights into the molecular virology and evolution of bornaviruses [52].

An increase in nucleotide sequence data has contributed to studies of novel exogenous viral nucleotide sequences and endogenous forms of viral elements. A well-known example of endogenous viral elements is the endogenous retroviruses (ERVs). During retrovirus replication, viral genomic RNA is reverse-transcribed and integrated into the host chromosome by viral reverse transcriptase (RT) and integrase. Therefore, once ERVs infect host germline cells and are integrated into the host chromosome, the integrated genomes are inherited by the host’s offspring as a part of their somatic genome [53]. Nearly one-tenth of the genomes of mammals is thought to be ERV elements, for example, approximately 8% of the human genome and 10% of the mouse genome are ERV sequences [54,55]. Although viruses do not leave physical fossils, endogenous forms of viruses are regarded as molecular fossils of past retroviral infections. Thus, ERVs provide valuable information on both ancient retroviruses and co-evolution of retroviruses and their hosts [53].

After ERVs were discovered in the late 1960s [56], non-retroviral RNA viruses were thought to be unable to integrate in their host genomes because they replicate without DNA intermediates. However, integration of non-retroviral RNA viruses has been reported occasionally [57–59]. In addition, since 2004, germline integration of non-retroviral RNA viruses has been reported in insects and plants [60–62]. However, no endogenous non-retroviral element was reported in the genomes of mammalian species before 2010.

2. Endogenous bornavirus-like elements

(a) Endogenous bornavirus-like elements in eukaryote genomes

During a study of the mechanism of BDV persistent infection, we unexpectedly found elements similar to BDV gene N in mammalian genomes, including the human genome, by BLAST analyses and designated them endogenous bornavirus-like N (EBLN) elements [51]. At approximately the same time, comprehensive analyses were carried out to detect non-retroviral virus-like elements in the genomes of eukaryotes and these studies found that there were also EBLNs in the genomes of other vertebrate animals [63,64]. Moreover, we have newly identified several EBLNs in vertebrate genomes, including snake, turtle, mole, afrotherians and primates (see electronic supplementary material, table S1).

Because synteny was not observed in these species, the EBLNs were probably not derived from a single integration event. However, several EBLNs were found to be orthologous in some species, which allowed their ages to be estimated [51,63,64] (see §2b).

In addition to EBLN, endogenous bornavirus-like M-, G- and L-like (EBLM, EBGL and EBLL, respectively) elements have been reported in many vertebrate genomes [63,64]. In addition, we have newly identified EBLM and EBLL elements in the genome of many eukaryotes, including invertebrates (see electronic supplementary material, tables S2 and S3). Overall, EBL elements have been identified in the genomes of many eukaryote species: primates, rodents, chiroptera, carnivore, afrotheria, insectivora, marsupials, reptiles, fishes, insect and spider.

Thus far, no element similar to X or P has been discovered, although X/P mRNA can integrate into the chromosomes of BDV-infected cells [65] (see §3a). Although this may be a coincidence, it is possible the X/P genes of exogenous viruses may be more mutable than other genes. Indeed, the amino acid sequences of X and P genes are less conserved than other genes in a comparison between BDV and ABV. For example, in BDV and ABV, the amino acid sequences of X and P are 40.7% and 59.7% identical, respectively, whereas the sequences of N are 72.5% identical [37]. Therefore, even if integration of an X/P segment had occurred, then its nucleotide sequence would have been quite different from the nucleotide sequences of modern bornaviruses, which would prevent its detection by BLAST analysis. However, because P was identified as a BDV pathogenic factor, expression of integrated P may have adverse effects on host animals. In fact, transgenic mice expressing BDV N protein do not show any sign of neuronal disease, whereas abnormal behaviour has been observed in BDV P transgenic mice [67,66]. Therefore, integration and expression of a P-like protein might have an adverse effect, producing a survival disadvantage for a host animal.

Among EBL elements, N- and L-derived elements are widely distributed in eukaryote genomes compared with other EBL elements, which may be due to the conservation of their amino acid sequences among bornaviruses. In addition, there is a transcription gradient from the 3' to 5' end of Mononegavirales genomes, resulting in an abundance of N mRNA [67]. Most cellular processed pseudo-genes, which are thought to be generated by the same mechanism as EBL elements, are derived from genes that are highly expressed in germline cells [68]. Therefore, the transcription gradient of bornaviruses may have contributed to the abundance of EBLN elements. However, although N and L mRNA of ancient bornaviruses might have been more susceptible to reverse transcription and integration, there is presently no known explanation for such processes.

(b) Paleovirology of bornaviruses

Because viruses do not leave traditional fossils, it is difficult to estimate the minimum ages of viruses. However, endogenous viral elements give us useful information for determining minimum ages, and we and others have identified several sets of orthologous EBL elements that enabled estimates of their minimum ages [51,63,64]. In the genomes of haplorhine primates, four copies of EBLNs, designated anthropoid EBLN-1 to -4, were shown to be orthologous, suggesting that integration of anthropoid EBLNs occurred in their common ancestor. Because the divergence of Strepsirrhini and Haplorhini is estimated to have occurred about 40 million years ago (Ma), the date at which endogenous anthropoid EBLNs and EBLG were established was at least 40 Ma. Because Afrotheria EBLNs in the African elephant (Loxodonta africana), Cape hyrax (Procavia capensis) and lesser hedgehog tenrec (Echinops telfairi) and their flanking sequences are readily aligned,
Afrotheria EBLNs seem to be orthologous, suggesting that they were established approximately 65 Ma [69]. Therefore, ancient bornaviruses may have coexisted with mammals and the history of bornavirus infection in mammals may go back to 65 Ma. This is the oldest endogenous non-retroviral RNA virus element identified so far. By contrast, integration of EBLN in *Ictidomys tridecemlineatus* (*I. tridecemlineatus* EBLN is designated itEBLN) is thought to have occurred recently (see §2d). Therefore, bornaviruses may have been coevolving with mammalian genomes for at least 65 Myr.

In addition, EBL elements are present in the genomes of many animals, suggesting that bornaviruses have infected such species of animals (see electronic supplementary material, tables S1–S3). Although such bornaviruses may not be easy to isolate, their nucleotide sequences may be detected in the near future. Therefore, the host range of bornaviruses may be much wider than previously thought.

(c) Establishment of endogenous bornaviruses

Bornaviruses do not encode RT and integrase genes, which leads to the question of how bornaviruses integrate into host genomes. In fact, there are several types of RT activity in cells, i.e. ERVs, retrotransposons and telomerase [70]. For example, an RNA segment of lymphocytic choriomeningitis virus (LCMV) is reverse-transcribed and integrated into its host cell chromosome by an intracisternal A-particle, anERV element in the mouse genome [57].

Several features of EBLNs suggest a possible mechanism for integration of ancient bornavirus-derived segments [51,63,64]. EBL elements are single gene-derived segments in diverse host genome locations, implying that the endogenous elements are derived from mRNAs of bornaviruses, not from bornavirus genomic or antigenic RNA segments. In addition, as shown in figure 2, each copy of anthropoid EBLN-1 to -4 contains a downstream 3' poly-A tail. In addition, direct repeat (DR) sequences flank *Homo sapiens* EBLNs (hsEBLNs). These data suggest that integration of bornavirus segments was mediated by long interspersed nuclear elements-1 (LINE-1).

LINE-1 is a retrotransposon and is widely distributed in the genomes of mammals. It inserts copies of itself by a copy and paste mechanism: LINE-1 is transcribed into mRNA and then reverse-transcribed and integrated into the cell genome by two proteins encoded by LINE-1 itself. LINE-1 is also involved in retrotransposition of short interspersed nuclear elements (SINEs) and in the formation of pseudo-genes by integration of cellular mRNA [71–73]. Integration mediated by LINE-1 produces an insertion with a 3' poly-A tail and flanked by a DR. Given the similar patterns, we suggest that integration of BDV segments was mediated by LINE-1. hsEBLN-2 may have integrated into the genome by recombination with the *Alu* element, which is a SINE (figure 2). The *Alu* element is immediately downstream of the hsEBLN-2 3' poly-A tail, with a nine nucleotide DR flanking the EBLN-2/*Alu* element. EBLNs in other species also have 3' poly-A tails [51,63,64]. However, there are EBLNs with neither a poly-A tail nor a DR. The accumulation of mutations may account for some elements lacking flanking DRs and/or a 3' poly-A tail. Some LINE-1-mediated insertions also do not have a 3'-poly-A tail or a DR [74–76]. It is interesting that the time when endogenous anthropoid EBLNs were established corresponds to the time at which LINE-1 actively retrotransposed in anthropoid genomes, supporting the LINE-1-mediated integration of anthropoid EBLNs model [77]. hsEBLN-1 to -4 also have DRs flanking the transcription start signal-like sequences and 3' poly-A tails. Therefore, anthropoid EBLN-1 to -4 seem to be generated by independent integration of full-length n gene mRNA.

LINE-1-mediated integration of bornavirus mRNA may have contributed to the evolution of mammalian genomes. In addition, LINE-1-mediated integration may have been involved in the integration of ancient filoviruses [63]. Because LINE-1 is active in many mammalian species, RNA viruses may still be involved in the evolution of mammals by integration of viral nucleotide segments.

(d) Biological significance of endogenous borna-like nucleoproteins

Several ERV-derived genes have a function in their host which is different than the function for which they were originally selected in the virus [53]. Fv1 and Fv4 are well-known examples and act as restriction factors of murine leukemia virus in mice. Such genes are called 'endogenous viral elements-derived immunity (EBI)' genes, i.e. they are genes with viral ancestry that act as inhibitors of viral infection [78]. Syncytin genes are derived from the *en* genes of ancient retroviruses and are essential for placenta formation in several mammalian species [79–82].

To consider whether EBL elements function like ERVs, we have focused on two EBLNs that have been relatively well studied: anthropoid EBLNs and itEBLN. Although many EBLNs became pseudo-genes due to premature stop codons and frameshifts, some anthropoid EBLNs retained relatively long open reading frames (ORFs); for example, hsEBLN-1 has an ORF encoding 366 amino acids, which is comparable with...
the full-length BDV N protein (370 amino acids). Similarly, there is a 272 amino acid ORF in hsEBLN-2 [51,63,64]. These ORFs are also conserved among other haplorhini primate species, although some of them are disrupted by insertion of a SINE. Because anthropoid EBLN-1 to -4 may have been acquired about 40 Ma, their ORFs may have been maintained for more than 40 Myr. In addition, hsEBLN-1 to -4 nucleotide sequences have been detected in the human EST database [63]. We also found that some EBLNs were expressed as mRNAs in several human and monkey cell lines. Furthermore, hsEBLN-2 was shown to be expressed as a protein in cultured cells and to interact with several human functional proteins [83]. These data suggested that some EBLNs encode functional proteins.

If anthropoid EBLNs encode functional proteins, then natural selection should have acted on them during primate evolution. Therefore, we examined whether natural selection had operated on anthropoid EBLNs by comparing synonymous and non-synonymous substitutions [84]. No evidence of natural selection acting on EBLN-1 to -4 during evolution of anthropoid species was found using this method of analysis. This suggested that current anthropoid EBLNs may not encode proteins whose function has been conserved among current anthropoid species or that EBLN products may not have had active functions from the divergence of Strepsirrhini and Haplorhini to the divergence of current species. In addition, our finding that extant EBLN-1 contains an ORF comparable with BDV N does not necessarily support the maintenance of N function after divergence of Strepsirrhini and Haplorhini [84]. However, our analyses do not account for stop codons produced by frameshifts and may not have been sensitive enough to detect past episodic natural selection. Further analyses are needed to determine whether EBLN products have had active functions during their evolution.

Even though EBLNs may not have been expressing functional proteins during their hosts' evolution, they may still have expressed a functionally relevant protein prior to divergence of Old and New World monkeys or were functioning as non-coding RNAs. EBLNs that have become pseudo-genes owing to elimination of the pathogenic viruses, leading to immunological tolerance to bornaviruses.

EBLNs may have encoded functional products before the divergence of Strepsirrhini and Haplorhini. For example, EBLNs may have prevented virus infection in their hosts early after the elements become endogenous, which might have provided selective advantages to the ancient primate hosts. This could have led to an increase of animals resistant to homologous or similar bornavirus infections followed by fixation of EBLNs in these species, thereby eliminating the bornavirus lineages. This loss could have resulted in the relaxation of constraints on EBLNs. Three possible mechanisms of EBLN activity as EDI have been considered. First, EBLNs may have altered the ratio of N to P gene products. It has been reported that a specific N:P ratio is required for BDV polymerase activity and that BDV N transgenic mice are resistant to BDV infection [66,88]. Therefore, expression of EBLNs may have inhibited the BDV polymerase by changing the N:P ratio. Second, EBLNs may have functioned as dominant negative proteins against the bornavirus N protein, like the ERV Gag protein that can block exogenous retroviruses in sheep at late stages of viral replication [89]. Finally, EBLNs may have been involved in immunity mediated by small RNAs as described above.

Alternatively, EBLNs may have been involved in immunological tolerance to ancient bornaviruses, conferring survival advantages to their hosts. However, this mechanism does not contribute to prevention of viral replication but to development of infection symptoms. Therefore, the host species may have become natural hosts of bornaviruses, or those bornavirus lineages may have been eliminated by other factors.

From these results, EBLNs, especially hsEBLN-2, may have had two different functions during evolution (figure 3). EBLNs may have acted as EDI genes early after becoming endogenous. The functional constrains on EBLNs may then have relaxed owing to elimination of the pathogenic viruses, leading to loss of the original EDI function. Finally, some EBLNs may have acquired novel functions. Current analyses of natural selection are not really designed to detect such patterns, which may explain why natural selection was not detected in our studies [84]. Further studies are needed to understand the biological significance of EBLNs. For example, because hsEBLNs are good models to study the function of EBLNs long after they became endogenous, it would be interesting to study the function of hsEBLN. In addition, itEBLN would be a good model for studying the function of EBLN soon after it became endogenous.

ItEBLN is thought to have become endogenous relatively recently [51,63,64]. However, phylogenetic analysis of EBLNs and N of exogenous bornaviruses showed that itEBLN is grouped in a cluster of exogenous bornaviruses. In addition, the amino acid sequence of itEBLN shows 77% identity with BDV N. Because there was no available genome sequence data for other squirrel species, we carried out Southern hybridization with genomic DNAs of several squirrel species [51]. These results suggested that itEBLN became endogenous after the divergence of Spermophilus and Marmota approximately 6 Ma [90,91]. We were not able to further analyse the age of itEBLN because of the lack of genomic DNA samples of closely related species of squirrels. A recent report suggested that LINE-1 in the I. tridecemlineatus genome is inactive and...
its retrotransposition activity may have ceased at 4–5 Ma. As described above, the 3’ poly-A tail is immediately downstream of itEBLN, suggesting involvement of LINE-1 in the generation of itEBLN [92]. It would be interesting to determine the age of itEBLN by genomic analyses of ground squirrels.

itEBLN also contains a long 203 amino acid ORF corresponding to a fragment of BDV N (residues 170–370). In addition, because the amino acid sequence identity of itEBLN and BDV N is very high, itEBLN and BDV N may have similar properties. These results suggested that itEBLN has a dominant negative effect on the BDV N protein. Further studies should investigate whether itEBLN affects BDV replication.

Considering the young age of itEBLN, this element might still be polymorphic in squirrel populations and play a role in antiviral defence against exogenous bornaviruses. Thus, it has been assumed that only squirrels having itEBLN are protected from exogenous bornavirus infection, as observed in the interaction between sheep and retroviruses [93]. Therefore, ground squirrels might be an extant model for elaborating the evolutionary relationships among exogenous RNA viruses, endogenous forms of these viruses and their hosts.

3. Integration of modern bornaviruses

(a) Integration of Borna disease virus mRNA in cultured cells

Discovery of endogenous bornavirus-like elements raised the question of whether segments of modern bornaviruses integrate into the host chromosome. We have performed a series of experiments with BDV showing that several BDV mRNAs (at least the mRNAs encoding N, X/P and N/X/P) are reverse-transcribed into cDNA in cultured BDV-infected cells [51] (figure 4a). N and X/P mRNAs have also been reported to be reverse-transcribed into cDNA in the brain of bank voles [65]. We also demonstrated that BDV N mRNA sequences are integrated into host genomic DNA in infected cultured cells and the mouse brain [51].

For insights into the integration of BDV mRNA, we tried to detect integrated BDV segments by novel Alu-PCR and inverse-PCR. In addition to primers in the N gene [51], we designed new primers in the P gene for inverse-PCR to detect insertions in N/X/P and X/P mRNAs. From 40 independent subclones of infected cells, 200 inverse-PCR reactions detected eight clones of integrated BDV mRNA, including the previously reported sequences in five clones [51]. As shown in figure 4b–d, we identified sequences of integrated N and X/P mRNAs. It is interesting that seven of the eight clones of integrated BDV segments have 3’ poly-A tails, similar to mammalian EBLNs. Truncation of the 5’ sequences of BDV mRNAs was also observed in several clones. Although no consensus host sequences for integration were found, the DRs derived from host sequences are located flanking some insertions. Deletions and rearrangements were also observed in the host genome sequences adjacent to integrated BDV segments. These results are similar to the features of LINE-1-mediated integration. Therefore, reverse transcription and integration of BDV mRNAs may also be mediated by LINE-1. Although we detected only N and X/P mRNAs, all BDV mRNAs have the potential to be recognized by the LINE-1 system because of the non-specificity of LINE-1.

(b) Possible novel pathogenicity of Borna disease virus

The fact that BDV mRNA has the potential to integrate into the host genome suggested that BDV might exhibit a novel pathogenicity [94,95], i.e. BDV might act as a mutagen in infected cells. Although LINE-1 is usually silenced by a host epigenetic mechanism in somatic cells, it has been reported that LINE-1 was active in human neural progenitor cells that were permissive for BDV infection [96,97].

In addition, BDV is a candidate RNA virus vector for gene therapy of CNS cells because, unlike other potential virus vectors, BDV establishes a long-term persistent infection [98]. RNA viruses were thought to be better vectors for gene therapy than DNA viruses, because RNA viruses were believed to carry no risk of integration into their target cell.
genomic DNA. Therefore, detailed studies of BDV integration will help elucidate both BDV pathogenicity and help determine the risk of using a BDV vector for gene therapy.

An important question is how often BDV integration occurs in infected cells. Although the integration efficiency of BDV mRNA was thought to be very low from inverse-PCR studies, the actual integration efficiency of BDV mRNA is still unclear. To evaluate BDV mRNA integration, we tried to estimate the integration efficiency of BDV N mRNA in BDV-infected cells by Alu-real-time PCR, which consists of two PCR steps, the first PCR with primers for Alu elements and the specific gene, and the second PCR with primers for the specific gene.

To generate control human genomic DNA, which should contain a randomly integrated single copy of a BDV N mRNA sequence per cell, we constructed a retrovirus vector containing the BDV N mRNA sequence and a puromycin resistance gene. OL (human oligodendrogliaoma cells) and

Figure 4. Reverse-transcription and integration of BDV mRNA. (a) Reverse transcription of BDV mRNAs and integrated BDV mRNAs. Viral genomic structure, transcripts, PCR amplified regions and PCR results are shown. PCR results are + (positive) or − (negative). (b) Structure of BDV N and X/P DNA. The mRNAs of N and X/P are BDV genome nucleotides 44 to 1193 and 1176 to 1883, respectively. The colours of the boxes correspond to those in (c,d). (c,d) Structure of integrated (c) BDV N and (d) X/P determined by inverse-PCR. Red boxes, integrated BDV N DNAs; blue boxes, BDV X/P DNAs; black boxes, poly-A sequences; grey boxes, truncated regions; red letters, direct repeat sequences; green letters, inserted nucleotides. Pre-integration genomic structure of chromosome 7 is shown in (c). Adapted from [51].
293 cells (human embryonic kidney cells) were infected with the retrovirus vector at a multiplicity of infection (m.o.i.) of 0.1 to ensure single virus infections. Two days after infection, cells were cultured with puromycin to eliminate cells in which the retrovirus vector did not integrate into the genomic DNA, followed by isolation of genomic DNA from these cells (figure 5a). Because a retrovirus typically produces one provirus [100], the genomic DNA should contain, at most, one DNA copy of viral N mRNA per cell. To make a series of standard genomic DNAs, genomic DNA containing DNA copies of the BDV N mRNA sequence (DNA/N mRNA) was serially diluted 10-fold with genomic DNA isolated from normal OL or 293 cells. We first evaluated the sensitivity of Alu-real-time PCR for BDV N mRNA by using the serially diluted genomic DNA/N mRNA, which showed that our Alu-real-time PCR system could detect one copy of integrated N mRNA per 1000 cells (data not shown).

To evaluate the infection efficiency of BDV N mRNA in persistently infected cells, we infected OL and 293 cells with BDV strain huP2br at an m.o.i. of 0.01. After infection, genomic DNA of OL and 293 cells infected with BDV (OL/BDV and 293/BDV, respectively) were collected every week for four weeks post-infection. At four weeks post-infection, approximately 100% and 80% of OL/BDV and 293/BDV cells, respectively, were infected (figure 5b). The integration efficiency of BDV-infected OL and 293 cells was analysed by Alu-real-time PCR, but the Alu-real-time PCR values were under the detection limit, which was one copy of integrated N mRNA per 1000 cells, as described above.

We also analysed the integration efficiencies of 20 subclones of each OL/BDV and 293/BDV cell by Alu-real-time PCR (figure 5c). These cells were maintained for at least eight weeks after the percentage of infected cells reached almost 100%. Relatively high integration efficiencies were found in seven subclones of OL/BDV, with the maximum in a subclone with one BDV copy per 148 cells. The high integration efficiencies were probably due to a bottleneck effect during subcloning. However, the integration efficiencies were lower than the detection limit in all of the 293/BDV and many of the OL/BDV subclone cells. The difference between OL and 293 may be due to the kinetics of viral replication. Alternatively, the LINE-1 activity may be different between these cells, although the RT activity was under the detection limit in both cell lines [51].

These data show that the BDV integration efficiency was much lower than that of retroviruses, even in these persistently infected cell lines. In this study, we were not able to determine the precise integration efficiency of BDV N mRNA owing to the detection limit. An improvement in sensitivity is required for determination of the integration efficiency. The integration efficiencies of other mRNAs have also not been determined. From the inverse-PCR results, the integration efficiency of X/P mRNA was comparable with or less than that of N mRNA. The integration efficiency of BDV mRNA probably correlates with both the amount of mRNA and the LINE-1 activity.

LINE-1-mediated integration theoretically could happen for all RNA viruses. In addition, like LCMV, other retroelements may cause the integration of RNA viruses [57]. Although the mechanisms are still unclear, several other RNA viruses have been reported to be integrated into their host chromosomes [58–62]. Therefore, RNA viruses have the potential for integration of virus-derived segments. Thus, analysis of integration events is important for understanding the pathogenicity of RNA viruses and virus vectors [51,57,95].

---

**Figure 5.**Integration of BDV N mRNA. (a) Schematic of the establishment of cell lines containing a single copy of the BDV N mRNA sequence. (b) Infection rates of BDV-infected OL and 293 cells as a function of time post-infection. (c) Integration efficiency of 20 subclones of each BDV-infected OL and 293 cell analysed by Alu-real-time PCR. n.d.; not detected.
4. Future

Since the discovery of EBL elements, it has been shown that non-retroviral RNA viruses and their vertebrate hosts have been associating and coevolving by integration of viral nucleotide sequences. Interestingly, these evolutionary processes may have resulted from the complex relationship among RNA viruses, retroelements and their host genomes. Thus far, there is no conclusion on whether EBL elements among RNA viruses, retroelements and their host genomes.

References


80. Goyiri G, Opolon P, Heidmann T. 2009 Specific murine envelope genes of retroviral origin are conserved in Muridae. *Cell*


