LETTERS

Endogenous non-retroviral RNA virus elements in mammalian genomes

Masayuki Horie[1]*, Tomoyuki Honda[1,2]*, Yoshiyuki Suzuki[3], Yuki Kobayashi[3], Takuji Daito[1], Tatsuo Oshida[4], Kazuyoshi Ikuta[1], Patric Jern[5], Takashi Gojobori[3], John M. Coffin[5] & Keizo Tomonaga[1,6]

Retroviruses are the only group of viruses known to have left a fossil record, in the form of endogenous proviruses, and approximately 8% of the human genome is made up of these elements[1,2]. Although many other viruses, including non-retroviral RNA viruses, are known to generate DNA forms of their own genomes during replication[3–5], none has been found as DNA in the germline of animals. Bornaviruses, a genus of non-segmented, negative-sense RNA virus, are unique among RNA viruses in that they establish persistent infection in the cell nucleus[6–8]. Here we show that elements homologous to the nucleoprotein (N) gene of bornavirus exist in the genomes of several mammalian species, including humans, non-human primates, rodents and elephants. These sequences have been designated endogenous Borna-like N (EBLN) elements. Some of the primate EBLNs contain an intact open reading frame (ORF) and are expressed as mRNA. Phylogenetic analyses showed that EBLNs seem to have been generated by different insertional events in each specific animal family. Furthermore, the EBLN of a ground squirrel was formed by a recent integration event, whereas those in primates must have been formed more than 40 million years ago. We also show that the *N* mRNA of a current mammalian bornavirus, Borna disease virus (BDV), can form EBLN-like elements in the genomes of persistently infected cultured cells. Our results provide the first evidence for endogenization of non-retroviral virus-derived elements in mammalian genomes and give novel insights not only into generation of endogenous elements, but also into a role of bornavirus as a source of genetic novelty in its host.

Bornaviruses are the only animal RNA viruses that achieve a highly cell-associated life cycle within the nuclear envelope[6–9], and can therefore provide not only new models of RNA virus replication, but also insight into dynamics of RNA molecules in eukaryote cells. In an effort to understand whether bornaviruses mimic host factors to maintain persistent infection in the nucleus, we searched human protein databases for sequences with similarity to BDV proteins. This search identified two hypothetical human proteins (GeneID LOC340900 and LOC55096), each of which has significant sequence similarity to BDV N (Fig. 1a and Supplementary Table 1). BDV N is a major structural protein, which tightly encapsidates the viral RNA to form the nucleocapsid. The LOC340900 sequence encodes a protein of comparable length (366 residues) to BDV N (370 residues), whereas LOC55096 seems to contain several frameshift mutations relative to BDV N, resulting in a shorter ORF length (Fig. 1a). Both LOC340900 and LOC55096 showed an overall 41% sequence identity and 58% similarity to BDV N and 72% identity to each other. The close relationship between BDV N and the homologous genes was further demonstrated by the alignment of transcription regulatory sequences on either side of BDV N (Fig. 1b). The S and T motifs in flanking sequences of both putative human proteins were well conserved with those of BDV (Fig. 1b). In addition, a poly-A sequence appears after the T1-like motif in the 3′ flanking region of LOC55096 (Fig. 1b). The homology of the human genes to BDV N was also confirmed by a permutation test (Supplementary Fig. 1). These findings indicated that both human genes may be endogenous elements related to BDV N gene, and therefore we designated them EBLNs (LOC340900, EBLN-1 and LOC55096, EBLN-2).

To investigate the presence of EBLN sequences in other animal species, we conducted tblastn searches using BDV N as a query in eukaryote and whole-genome shotgun databases at NCBI. Sequences with blast *E*-values of 10^{-10} or lower were identified as EBLNs. We found two additional human elements (EBLN-3 and -4) as well as a number of related sequences in various mammalian species, including marsupials (Supplementary Table 2). Orthologous genes to human EBLNs were identified in the genomes of non-human anthropoid primates, including chimpanzee, gorilla, orang-utan, and macaque (Supplementary Table 2). We also detected primate EBLNs in the genomes of the suborder Strepsirrhini, including the mouse lemur and Garnett's galago. Furthermore, two species of the Afrotheria, African elephant and cape hyrax, and four rodents were found to have EBLNs with *E*-values of less than 10^{-20} (Supplementary Table 2). An EBLN locus with a high level of similarity to BDV N was also identified in the thirteen-lined ground squirrel (TLS) genome (Supplementary Fig. 2a). Like the human EBLNs, the TLS EBLN contained a 3′ poly-A sequence, as well as S and T signal motifs, in its 3′ flanking region (Supplementary Fig. 2b). Almost all EBLN fragments, except for EBLN-1 and the TLS gene, contained several stop codons in the predicted coding sequences, or lacked the identifiable flanking sequences. In addition, we found that all anthropoid EBLNs, except for EBLN-4, are expressed as mRNAs in some human and monkey-derived cell lines (Supplementary Fig. 3). A previous study reported the interaction of human EBLN-2 with other cellular proteins, such as AP1S1, TUSC2/FUS1 and FANCC (ref. 10) (Supplementary Table 1), indicating that anthropoid EBLNs may encode functional proteins.

To investigate whether other mammalian species contain EBLN-related sequences in their genomes further, we conducted Southern blot hybridization under low-stringency conditions using human, murine and TLS EBLN as probes (Fig. 1c and d). Along with the clear signals in primate genomes, we detected reproducible faint positive bands in murine and shrew genomes when using a human EBLN probe (Fig. 1c, dots). The signals were also observed using a mouse

[1]Department of Virology, Research Institute for Microbial Diseases (BIKEN), Osaka University, Osaka 565-0871, Japan. [2]Japan Society for the Promotion of Science (JSPS), Chiyoda-ku, Tokyo 102-8472, Japan. [3]Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan. [4]Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan. [5]Department of Molecular Biology and Microbiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA. [6]PRESTO, Japan Science and Technology Agency (JST), Chiyoda-ku, Tokyo 102-0075, Japan.
*These authors contributed equally to this work.

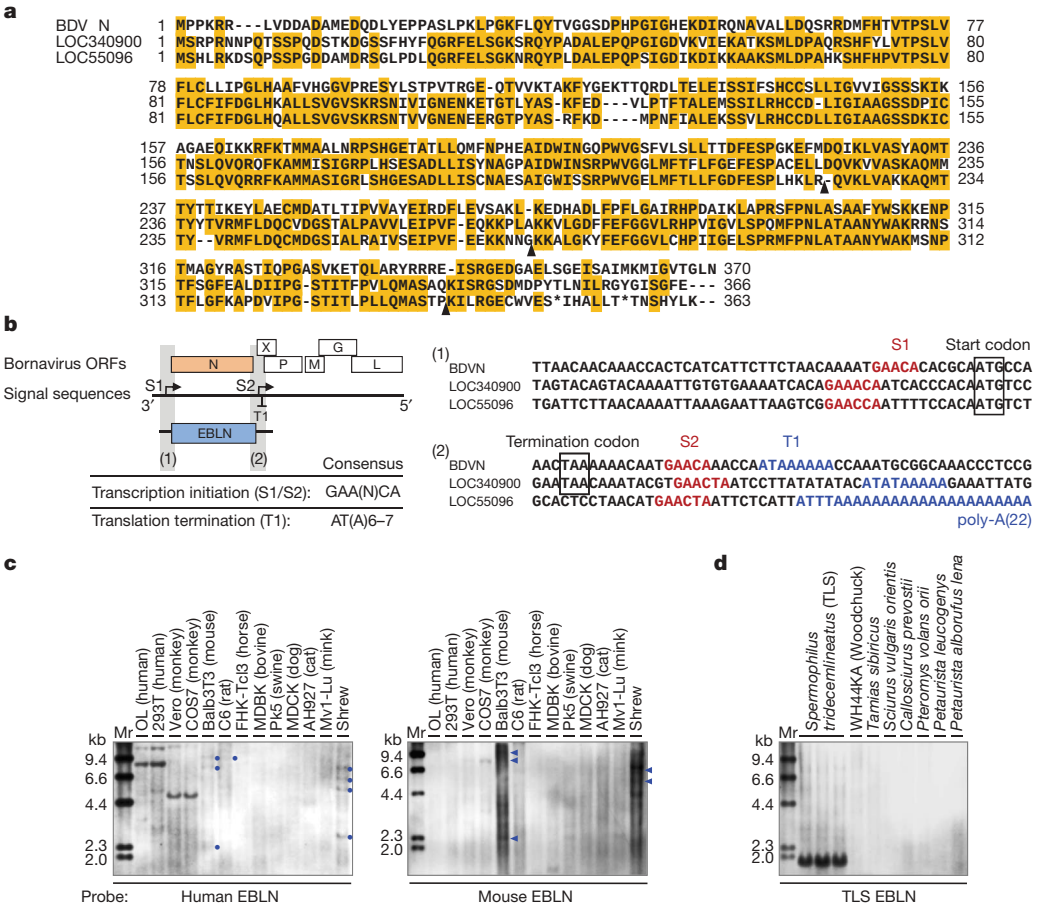

Figure 1 | Bornavirus N-like elements in mammalian genomes. **a**, Alignment between predicted amino acid sequences of BDV N and two human bornavirus N-like elements. Black arrowheads indicate predicted frameshift sites in LOC55096. **b**, Sequence alignments of transcription signal sites (S1/S2 and T1) at both the 5′ and 3′ ends of the bornavirus N ORF. A schematic representation of bornavirus genome structure is shown. **c**, **d**, Low-stringency Southern blot hybridizations of DNA from various mammalian species using human EBLN-1 and mouse EBLN chr.11 (**c**) and TLS EBLN (**d**) as probes. Dots and arrowheads on the right side of the murine and shrew lanes in panel **c** indicate the positions of reproducible positive signals. Mr, Molecular marker.

EBLN probe (Fig. 1c, arrowheads), indicating that the faint bands are most likely to be EBLN-related sequences. In fact, EBLN-like sequences, albeit with *E*-values greater than 10^{-10}, were found in the Eurasian shrew genome in our tblastn searches. On the other hand, except for TLS, no positive band was detected by the TLS probe in the genomes of several different squirrel species, such as woodchuck (*Marmota* spp.), the closest species to the TLS (*Spermophilus* spp.) (Fig. 1d)[11], indicating that the ground squirrels are likely to be the only host species of EBLN within the squirrel family. The BDV N probe detected many faint and smear bands that include the signals detected by EBLN-specific probes in both selected mammalian species and the squirrel families (Supplementary Fig. 4), indicating that EBLN-related fragments are more widely distributed in the mammalian genome.

We next performed a comprehensive phylogenetic analysis using nucleotide sequences of all EBLNs with *E*-values less than 10^{-20} (Fig. 2 and Supplementary Fig. 5). In addition to EBLNs, we included avian bornaviruses (ABVs)[12] and an exogenous reptile bornavirus (RBV) sequence, which was detected in a cDNA library from a *Bitis gabonica* (Gaboon viper) venom gland[13] (Supplementary Fig. 6). As shown in Fig. 2, the anthropoid and murine EBLNs are clustered phylogenetically within each host order. By contrast, EBLNs from other species, including African elephant, cape hyrax and guinea pig, form branches independent from the evolutionary lineage of their hosts, indicating that these EBLNs had most likely invaded each species via independent integration events. Interestingly, the TLS EBLNs form a tight cluster more closely related to modern exogenous bornaviruses than to those of other animals. Considering that a closely related species does not contain EBLNs, the integration of squirrel EBLN could have been a very recent event. A phylogenetic analysis using all primate EBLNs, including marmoset (Supplementary Fig. 7), showed that the integration events leading to the primate EBLNs occurred in the Haplorrhini at least before the split between rhesus macaque and marmoset.

To investigate whether current bornaviruses are able to be copied into DNA to produce EBLN-like elements, we first performed PCR analyses using DNA of persistently BDV-infected cells. As shown in Fig. 3a and Supplementary Table 3, BDV DNA was clearly detected in some cell lines by a primer set targeted to the BDV N region. To understand which viral RNA species serve as template for the DNA form of BDV, we used several primers within the BDV genome for amplification. The results showed that primer sets straddling the boundaries of BDV transcription units could not amplify BDV-specific DNA (Fig. 3b and c), indicating that the DNA is transcribed from mRNAs of BDV. We detected BDV-specific DNA in the brains of persistently BDV-infected mice (Supplementary Fig. 8), indicating that BDV can produce DNA forms *in vitro* and *in vivo*. We next performed Alu-PCR to investigate whether BDV DNA detected in the infected cells exists as integrated or extrachromosomal DNA. As shown in Fig. 3d and Supplementary Fig. 9, an Alu-specific PCR product was detected in BDV-infected cells only when using an N-specific forward primer about 30 days post-infection. This observation indicated that although BDV DNA in infected cells may be mainly extrachromosomal, the N gene is integrated into the host genome during persistent infection.

We further characterized the BDV DNA insertions and flanking cellular sequences by using Alu-PCR and inverse PCR (Supplementary Fig. 10)[14]. Integration sites were present on various chromosomes

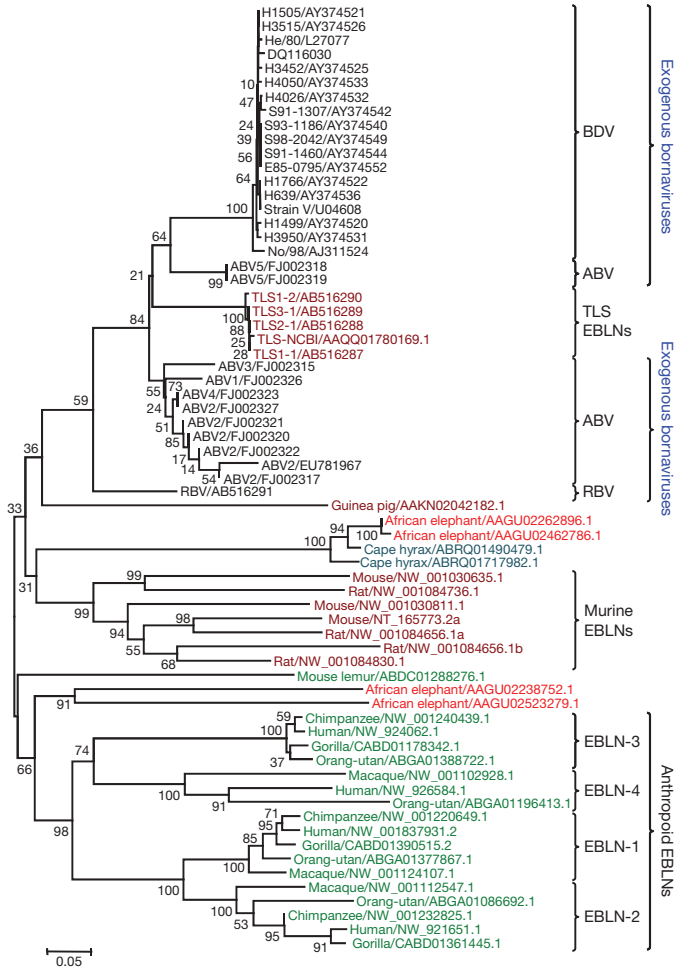

Figure 2 | Phylogenetic tree of exogenous bornaviruses and mammalian EBLNs. The bootstrap probability is indicated for each interior branch. The scale bar indicates the number of amino acid substitutions per site. Animals belonging to the same order are indicated by the same colour. Strain and sequence accession numbers are given for each sequence.

(Fig. 4). Similar to some mammalian EBLNs, many BDV DNA insertions contained a 3′ poly-A sequence (Fig. 4b and c). In addition, integrations of truncated BDV N DNA were also found in some clones. No apparent consensus sequences were found at the sites, although target site duplications (TSDs) were detected in some clones from the inverse PCR (Fig. 4c). We also found deletions, as well as sequence rearrangement, of host genome adjacent to BDV DNA insertions (Fig. 4c). These results indicate that modern BDV is able to produce DNA forms leading to insertion of EBLN-like elements into its host's genome.

This report is the first to provide evidence of endogenous sequences derived from a non-retroviral RNA virus in mammalian species. Phylogenetic analyses demonstrate that the oldest primate EBLN observed must have appeared in an ancestor of primates after the separation between Strepsirrhini and Haplorrhini, implying that bornaviruses have coexisted with primates for an evolutionary history stretching at least 40 million years. Thus, bornaviruses are the first non-retroviral RNA virus whose existence in prehistoric times has been confirmed. To date, the evolution/origin of RNA viruses is a major puzzle in the relationship between viruses and mammalian hosts, because simple molecular clock calculations using an average rate of nucleotide substitutions estimate the origin of RNA viruses to be a very recent event[15–17]. Despite replication during tens of millions of years as exogenous viruses, the amino acid sequences of current BDV N seem surprisingly conserved relative to EBLNs. This conservation demonstrates the inapplicability of simple molecular clocks to RNA virus evolution. Discovery of EBLNs in several

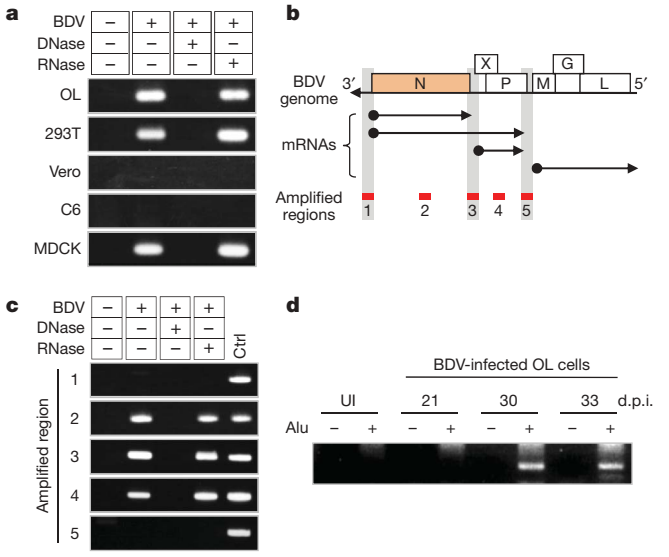

Figure 3 | Reverse transcription and integration of BDV RNA in mammalian cells. **a**, PCR amplification of BDV N-specific cDNA in BDV-infected cells. OL and 293T, human; Vero, monkey; C6, rat; MDCK, dog. **b**, Schematic representation of the bornavirus genome and mRNAs. Regions for the PCR amplification are indicated by red bars. **c**, Region-dependent amplification of BDV cDNA in infected OL cells. The numbers on the left side of the panels correspond to the amplification regions in panel **b**. Ctrl indicates the results of RT–PCR using RNA from BDV-infected OL cells. **d**, Integration of BDV DNA. Genomic DNA was isolated from BDV-infected OL cells at the indicated days after infection, and Alu-PCR was performed with (+) or without (–) the Alu primer. UI, genomic DNA from uninfected cells; d.p.i., days post-infection.

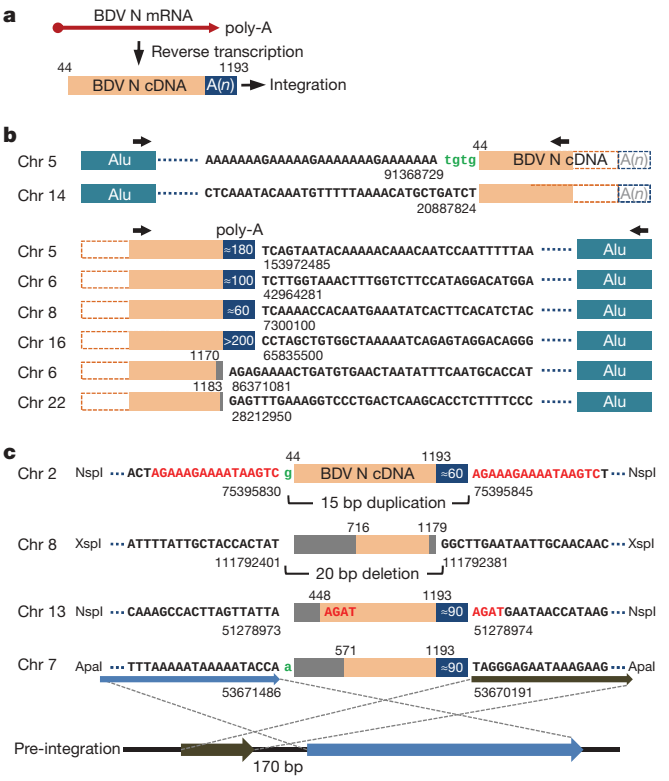

Figure 4 | Structures of BDV N integration events in OL cells. **a**, Structure of BDV N cDNA. The numbering corresponds to nucleotide positions in the BDV genome. The BDV N transcript runs from nucleotide positions 44 to 1193. **b**, **c**, Structures of BDV N integrations detected by Alu-PCR (**b**) and inverse PCR (**c**). Grey rectangles in the N cDNA indicate truncated regions. Black lettering, host genome sequences; green lettering, inserted nucleotides; red lettering, predicted TSDs. The blue box indicates the position and length of the poly-A sequence. The pre-integration form of chromosome 7 is shown in panel **c**.

mammalian species will help shed light on the evolutional history of RNA viruses and their hosts.

The sequence characteristics of both EBLNs and BDV DNA insertions in host genomes indicate that the reverse transcriptase activity encoded by retrotransposons, such as long interspersed nucleotide elements (LINEs), is likely to be involved in the reverse transcription and integration of bornavirus mRNAs, although some clones showed no apparent TSDs (ref. 18). LINE-1s (L1) are abundant retrotransposons, whose enzymes are able to sometimes target cellular mRNAs and produce processed pseudogenes in mammalian genomes[19–21]. The organization of sequences flanking EBLN-2 is consistent with the action of L1. The sequence shows the presence of an AluSx element immediately downstream of the 3′ poly-A tail of EBLN-2 (Supplementary Fig. 11). The key observation is that the EBLN-2/AluSx element is flanked by a perfect 9-bp TSD. Because the AluSx itself is not flanked by TSDs and the 3′ end of Alu is known to be recognized by L1 during target-primed reverse transcription, the presumed EBLN-2/AluSx chimaera element was most likely created and integrated by the L1 machinery. Thus, it is likely that EBLNs are processed pseudogenes derived from ancient bornavirus infections. At present, the reasons why bornaviruses but not other non-retroviral RNA viruses, and why only N and not other genes, have been preserved in mammalian genomes as endogenous elements are not clear. There are several possibilities. First, bornaviruses may have greater access to the germline. Second, the BDV N mRNA, like some cellular RNAs, may have features that, by chance, make it a favourable template for L1-mediated reverse transcription[22,23]. Third, the predominant transcription of BDV N mRNA in infected cells may also favour its association with the L1 replication machinery. The selectivity for BDV N mRNA implies a role for specific structural features, perhaps in conjunction with one or more of the other possibilities. Our data also raise the possibility that, like some endogenous retroviruses, EBLNs may have some function in their host species. An analysis of the non-synonymous to synonymous substitution ratios among anthropoid EBLNs indicates functional, albeit weak, evolutionary conservation. This finding implicates bornaviruses as a new source of genetic innovation in their hosts. Further studies will be needed to explore this possibility.

METHODS SUMMARY

Homology searches (blastp, tblastn) were conducted using the amino acid sequence of BDV N H1499 (International Nucleotide Sequence Database accession number AY374520) as a query and the genomic sequences of 234 eukaryotes as a database at the genomic blast server at the National Center for Biotechnology and Information, NCBI. Sequence hits with *E*-values less than 10^{-10} were collected together with neighbouring hits, if any, with higher *E*-values and combined according to their alignment pattern with BDV N. The resulting amino acid sequence was examined for the presence of a BDV_P40 domain (Pfam accession number PF06407.3) using HMMPFAM. The sequence was identified as a putative EBLN when the domain was detected with the *E*-value of less than 10^{-10}.

The putative EBLN amino acid sequences that were identified with *E*-value of less than 10^{-20} in both tblastn and HMMPFAM were used for the phylogenetic analysis with N sequences of various exogenous bornaviruses. The multiple alignments of EBLN and BDV N amino acid sequences were made according to the alignment pattern of EBLN sequences to BDV N in the tblastn results. The phylogenetic tree was constructed using the neighbour-joining method[24] and the evolutionary distance measured as the proportion of difference (*p* distance) with the pairwise deletion option in MEGA (version 4.0)[25]. The reliability of interior branches in the phylogenetic tree was assessed by the bootstrap method with 1,000 resamplings.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 2 September; accepted 17 November 2009.

1. Jern, P. & Coffin, J. M. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42,** 709–732 (2008).
2. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
3. Zhdanov, V. M. Integration of viral genomes. *Nature* **256,** 471–473 (1975).
4. Klenerman, P., Hengartner, H. & Zinkernagel, R. M. A non-retroviral RNA virus persists in DNA form. *Nature* **390,** 298–301 (1997).
5. Geuking, M. B. *et al.* Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* **323,** 393–396 (2009).
6. Tomonaga, K., Kobayashi, T. & Ikuta, K. Molecular and cellular biology of Borna disease virus infection. *Microbes Infect.* **4,** 491–500 (2002).
7. de la Torre, J. C. Molecular biology of Borna disease virus and persistence. *Front. Biosci.* **7,** d569–d579 (2002).
8. Lipkin, W. I. & Briese, T. in *Fields Virology* 5th edn (eds Knipe, D. M. & Howley, P. M.) 1829–1851 (Lippincott Williams & Wilkins, 2007).
9. Chase, G. *et al.* Borna disease virus matrix protein is an integral component of the viral ribonucleoprotein complex that does not interfere with polymerase activity. *J. Virol.* **81,** 743–749 (2007).
10. Ewing, R. M. *et al.* Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Syst. Biol.* **3,** 89 (2007).
11. Mercer, J. M. & Roth, V. L. The effects of Cenozoic global change on squirrel phylogeny. *Science* **299,** 1568–1572 (2003).
12. Kistler, A. L. *et al.* Recovery of divergent avian bornaviruses from cases of proventricular dilatation disease: identification of a candidate etiologic agent. *Virol. J.* **5,** 88 (2008).
13. Francischetti, I. M., My-Pham, V., Harrison, J., Garfield, M. K. & Ribeiro, J. M. Bitis gabonica (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* **337,** 55–69 (2004).
14. Hui, E. K., Wang, P. C. & Lo, S. J. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. *Cell. Mol. Life Sci.* **54,** 1403–1411 (1998).
15. Holmes, E. C. Molecular clocks and the puzzle of RNA virus origins. *J. Virol.* **77,** 3893–3897 (2003).
16. Duffy, S., Shackelton, L. A. & Holmes, E. C. Rates of evolutionary change in viruses: patterns and determinants. *Nature Rev. Genet.* **9,** 267–276 (2008).
17. Korber, B., Theiler, J. & Wolinsky, S. Limitations of a molecular clock applied to considerations of the origin of HIV-1. *Science* **280,** 1868–1871 (1998).
18. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31,** 159–165 (2002).
19. Maestre, J., Tchenio, T., Dhellin, O. & Heidmann, T. mRNA retroposition in human cells: processed pseudogene formation. *EMBO J.* **14,** 6333–6338 (1995).
20. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nature Genet.* **24,** 363–367 (2000).
21. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
22. Zhang, Z., Carriero, N. & Gerstein, M. Comparative analysis of processed pseudogenes in the mouse and human genomes. *Trends Genet.* **20,** 62–67 (2004).
23. Pavlicek, A. & Jurka, J. in *Genomic disorders* (eds Lupski, J. R. & Stankiewicz, P.) 57–72 (Humana Press, 2006).
24. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4,** 406–425 (1987).
25. Tamura, K., Dudley, J., Nei, M. & Kumar, S. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24,** 1596–1599 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kawahara for helping the capture of the wild shrews (*Sorex unguiculatus* and *Sorex gracillimus*) at Kiritappu wetland, Hokkaido, Japan. We thank I. Francischetti for provision of Gaboon viper (*Bitis gabonica)* venom gland tissue and a cDNA library, D. Vaughan for thirteen-lined ground squirrel (*Spermophilus tridecemlineatus*) brain and liver tissues, and K. Maeda, T. Miyazawa and N. Ohtaki for providing culture cell lines from several mammalian species. This work was supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grants-in-aid for Scientific Research on Priority Areas (Infection and Host Responses; Matrix of Infection Phenomena) (K.T.), PRESTO (RNA and Biofunctions) from Japan Science and Technology Agency (JST) (K.T.), a Health Labour Sciences Research Grants for Research on Measures for Intractable Diseases (H20 nanchi ippan 035) from the Ministry of Health, Labor and Welfare of Japan (K.T.), research grant R37 CA 089441 from the National Cancer Institute (J.M.C.) and a fellowship from the Wenner-Gren Foundation (P.J.). J.M.C. was a Research Professor of the American Cancer Society with support from the George Kirby Foundation.

Author Contributions K.T. designed research; M.H., T.H., T.D. and K.T. conducted experiments using virus and culture systems; T.O. collected samples; Y.S., Y.K. and T.G. performed phylogenetic analysis; M.H., T.H., Y.S., K.I., P.J., T.G., J.M.C. and K.T. analysed data; and M.H., Y.S., P.J., J.M.C. and K.T. wrote the manuscript. All authors discussed the results.

Author Information The TLS EBLN and RBV sequences reported here have been deposited in the DDBJ/EMBL/GenBank and the accession numbers are shown in Figure 2. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.T. (tomonaga@biken.osaka-u.ac.jp).

METHODS

Permutation test. A permutation test was conducted to examine the homology of human EBLNs to the N gene of BDV, taking into account their base composition. The nucleotide sequence of each EBLN was aligned with that of the BDV N gene (strain CRP3A: accession number AY114161) using CLUSTAL W. Gaps were eliminated from the alignment, and the proportion of identical sites (q) was computed. Nucleotide sequences of both the EBLN and the BDV N gene were randomly permuted using pseudorandom numbers, and the q value was computed as indicated above. The permutation process was repeated 10,000 times, and the distribution of the q value between two unrelated sequences of the same base composition as the original EBLN and the N gene was obtained. The probability (p) of observing the q value equal to or greater than the original value in the comparison of unrelated sequences was obtained from the distribution.

Tissue samples. Tissues from three weanling thirteen-lined ground squirrel (*Spermophilus tridecemlineatus*) born in May 2008 (four generations from wild stock) were provided from the Ground Squirrel Captive Breeding Colony at the University of Wisconsin Oshkosh, USA. Immediately after decapitation, brain and liver were rapidly dissected, cut into 5 mm cubes, immersed in chilled methanol, and stored frozen in liquid nitrogen until use. Shrew tissues (brain and liver) were isolated from wild-captured long-clawed shrews (*Sorex unguiculatus*) in Hokkaido, Japan. The shrews were captured under sampling permission of the government of Hokkaido. Immediately after capture, tissue samples were fixed in RNAlater (Ambion) and stored frozen until use. Gaboon viper (*Bitis gabonica*) venom gland tissue was obtained as frozen samples from the Laboratory of Malaria and Vector Research at National Institute of Allergy and Infectious Diseases, National Institutes of Health, USA. Ethanol-fixed tissues from Siberian flying squirrels (*Pteromys volans orii*) and Eurasian red squirrels (*Sciurus vulgaris orientis*) were obtained from the Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan.

DNA isolation. Total DNA from cultured cells was isolated using QIAamp DNA Blood Mini kit (Qiagen). Two monkey cell lines, Vero and COS7, used in this study are derived from African green monkey. High molecular mass DNA was extracted by using a Blood and Cell Culture DNA Mini kit (Qiagen). Genomic DNAs of shrews, ground squirrels and the Gaboon viper were prepared from tissue samples using a phenol/chloroform extraction method or the Blood and Cell Culture DNA Mini kit. To minimize the risks of contamination, DNA extraction was performed in UV-irradiated safety cabinet with UV-irradiated pipettes, tubes and filter tips.

DNA samples. Genomic DNAs from chipmunks (*Tamias sibiricus*), Japanese giant flying squirrels (*Petaurista leucogenys*) and red and white giant flying squirrels (*Petaurista alborufus lena*) were obtained from the Department of Life Science and Agriculture, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido, Japan.

Southern blot hybridization. Genomic DNA (5 μg) was digested with appropriate restriction endonucleases (TaKaRa). After electrophoresis in a 0.9% agarose gel, DNA was transferred onto positively charged Nylon membranes (Roche) and baked at 120 °C for 30 min. The membrane was prehybridized in DIG Easy Hyb (Roche) at 32 °C for 30 min. Human and TLS EBLN and BDV N probes were labelled by DIG-High Prime (Roche). Hybridization was performed in DIG Easy Hyb containing 25 ng ml^{-1} probe at 32 °C overnight. The membrane was washed twice with 2× SSC, 0.1% SDS at room temperature for 5 min, and then washed twice with 0.5× SSC, 0.1% SDS at 50 °C for 15 min. For chemiluminescence detection, Anti-DIG-alkaline phosphatase, Fab (Roche) and CDP-Star (Roche) were used according to the manufacturer's instructions. The low-stringency condition can theoretically detect sequences having at least 75% identity with each probe.

F-PERT assay. F-PERT (fluorescent product-enhanced reverse transcriptase) assay was performed as described previously[26]. Briefly, cells were lysed in disruption buffer (40 mM Tris-HCl, pH 8.1; 50 mM KCl; 20 mM dithiothreitol; 0.2% NP-40) and the protein concentration was measured. For the reverse transcription reaction, 1 μg of the cellular protein in 10 μl disruption buffer and an equal volume of 2× RT mix (100 mM KCl; 20 mM Tris-HCl pH 8.3; 11 mM $MgCl_2$; 1 mM dATP, dCTP, dGTP and dTTP; 0.4 μM reverse primer: 5′-CACAGGTCAAACCTCCTAG GAATG-3′, 0.2% NP-40; 20 mM dithiothreitol; 0.8 U μl^{-1} RNasin (Promega); 314 ng μl^{-1} calf thymus DNA (Sigma) and 1.5 ng MS2 RNA (Roche)) were mixed and incubated at 48 °C for 30 min. cDNA was mixed with forward primer: 5′-TCCTGCTCAACTTCCTGTCGAG-3′, reverse primer, probe: 5′-(FAM)-TC TTTAGCGAGACGCTACCATGGCTA-(TAMRA)-3′ and 2× TaqMan Universal PCR Master Mix (Applied Biosystems). Real-time PCR was carried out in an ABI 7900HT Fast Real-Time PCR System using the following parameters: 95 °C 10 min, then 50 cycles consisting of 94 °C for 30 s and 64 °C for 1 min. SuperScript III reverse transcriptase (Invitrogen) was used as standard control.

Virus infection. The BDV strains, huP2br, He/80 and recombinant BDV expressing GFP (rBDV-5′ GFP), were used in this study. Virus stock was prepared from the supernatants of BDV-infected cells. Confluent BDV-infected cells were washed with 20 mM HEPES, pH 7.5 and incubated with 5 ml of 20 mM HEPES (pH 7.5) containing 250 mM $MgCl_2$ and 1% FCS for 1.5 h at 37 °C. Supernatants were harvested and centrifuged at 2,500g for 5 min. The resulting supernatants were used for virus stock. The infectious titre was determined by focus forming assay as described previously[27]. The cell lines used in this study were cultured in Dulbecco's modified Eagle's medium (DMEM)-containing 10% fetal bovine serum (FBS). Newborn Balb/c mice (Oriental kobo) were inoculated intracranially with 200 focus forming units of BDV stock per animal within 24 h after birth. Infected animals were sacrificed at 21 days post-infection. The brains were collected for further analyses. All animal experiments conformed to the guide for the care and use of laboratory animals in the Research Institute for Microbial Diseases, Osaka University, Japan.

Alu-PCR analysis. Integration of BDV sequences into host genomes was detected by using primers specific to human Alu repeats and to BDV N region. First round amplification was performed in a final volume of 25 μl containing 0.5 U Ex Taq (TAKARA), 1× Ex Taq buffer, 0.2 mM dNTP, BDV N-specific primer, Alu primer and 100 ng of high molecular mass genome DNA. As control, PCR without the Alu primer was also performed. The condition of first PCR was as follows: denature for 5 min, 20 cycles of 94 °C for 30 s, 53 °C for 30 s, 72 °C for 4 min, followed by an extended elongation at 72 °C for 10 min. The second round PCR reaction was carried out with 1 μl of the first reaction using BDV N-specific nested primers. The reaction was run as follows: denature for 5 min, 40 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 20 s with the final extension at 72 °C for 3 min. The sequence information for primers used in Alu-PCR is available on request.

Amplification of virus–host junction. Virus–host junctions were amplified by using Alu-PCR and inverse PCR methods. Alu-PCR analysis was performed as described previously[28]. Briefly, the first round PCR reaction was carried out with 100 ng of high molecular mass genome DNA in a final volume of 25 μl containing 0.5 U Ex Taq, 0.2 mM dNTP, 2 μM BDV-specific primer and 0.2 μM Alu primer under the following conditions: denaturing at 94 °C for 1 min, 10 cycles of 94 °C for 30 s, 59 °C for 30 s, 70 °C for 3 min, followed by an extended elongation at 70 °C for 10 min. After amplification, 0.5 U of uracil DNA glycosylase (New England Biolabs) was added into the tubes and incubated at 37 °C for 30 min. After heating at 94 °C for 10 min to break DNA strands at apurinic dUTP sites, the next amplification primers, Tag- and BDV-specific primers, were added. Second round PCR was performed as follows: after denaturing at 94 °C for 2 min, 20 cycles of touchdown PCR in which the annealing temperature was decreased one degree every other cycle from 65 °C to 56 °C. The remaining 20 cycles were run with the annealing temperature at 55 °C, followed by an extended elongation at 72 °C for 3 min. One microlitre of the second round PCR products was further amplified with Tag- and BDV-specific primers as follows: after denaturing for 2 min, 25 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 3 min with the final extension at 72 °C for 3 min. Amplified DNA was electrophoresed, extracted and then sequenced.

Inverse PCR was described elsewhere[29]. Briefly, 1 μg genomic DNA was digested with an appropriate restriction enzyme, including ApaI, BamHI, EcoRI, NspI, PstI or XspI, for 3 h. Digested DNA was purified with QIAquick PCR Purification kit (Qiagen) and diluted with T4 DNA ligase buffer to a final DNA concentration of 1 ng μl^{-1}, and then T4 DNA ligase (New England Biolabs) was added to a final concentration of 4 U μl^{-1}. After ligation at 16 °C for 16 h, ligated DNA was isolated using a QIAquick PCR Purification kit. Five microlitres of the eluate were used for nested PCR. First round PCR was conducted in a 50 μl final volume containing 1 U TaKaRa Ex Taq, 0.2 mM dNTP and 0.2 μM BDV-specific primer set with the following program: after denaturing at 94 °C for 2 min, 20 cycles of 94 °C for 30 s, 70 °C for 30 s (temperature was decreased one degree every other cycle), 72 °C for 4 min and 20 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 4 min with the final extension at 72 °C for 3 min. Second round PCR was performed with 1 μl of the first reaction. The reaction condition was 94 °C for 2 min, 25 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 4 min with the final extension at 72 °C for 3 min. PCR products were electrophoresed and DNA was extracted from the desired bands and sequenced. Sequence information for primers used in this study is available on request.

26. Lovatt, A. *et al.* High throughput detection of retrovirus-associated reverse transcriptase using an improved fluorescent product enhanced reverse transcriptase assay and its comparison to conventional detection methods. *J. Virol. Methods* **82**, 185–200 (1999).
27. Ohtaki, N. *et al.* Downregulation of an astrocyte-derived inflammatory protein, S100B, reduces vascular inflammatory responses in brains persistently infected with Borna disease virus. *J. Virol.* **81**, 5940–5948 (2007).
28. Minami, M., Poussin, K., Brechot, C. & Paterlini, P. A novel PCR technique using Alu-specific primers to identify unknown flanking sequences from the human genome. *Genomics* **29**, 403–408 (1995).
29. Wo, Y. Y., Peng, S. H. & Pan, F. M. Enrichment of circularized target DNA by inverse polymerase chain reaction. *Anal. Biochem.* **358**, 149–151 (2006).