The Journal of Veterinary Medical Science

FULL PAPER

Virology

Systematic estimation of insertion dates of endogenous bornavirus-like elements in vesper bats

Yahiro MUKAI[1,2)], Masayuki HORIE[1,3)]* and Keizo TOMONAGA[1,2,4)]

[1)]Laboratory of RNA Viruses, Department of Virus Research, Institute for Frontier Life and Medical Sciences (InFRONT), Kyoto University, Kyoto 606-8507, Japan
[2)]Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto 606-8501, Japan
[3)]Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8501, Japan
[4)]Department of Molecular Virology, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan

J. Vet. Med. Sci.
80(8): 1356–1363, 2018
doi: 10.1292/jvms.18-0211

Received: 13 April 2018
Accepted: 11 June 2018
Published online in J-STAGE: 5 July 2018

ABSTRACT. Endogenous bornavirus-like elements (EBLs) are sequences derived from bornaviruses (the family *Bornaviridae*) that are integrated into animal genomes. They are formed through germline insertions of segments of bornaviral transcripts into animal genomes. Because EBLs are molecular fossils of bornaviruses, they serve as precious sources of information to understand the evolutionary history of bornaviruses. Previous studies revealed the presence of many EBLs in bat genomes, especially in vesper bats, and suggested the long-term association between bats and bornaviruses. However, insertion dates of EBLs are largely unknown because of the limitations of available bat genome sequences in the public database. In this study, through a combination of database searches, PCR, and sequencing approaches, we systematically determined the gene orthologies of 13 lineages of EBLs in bats of the genus *Myotis* and *Eptesicus* and family Vespertilionidae. Using the above data, we estimated their insertion dates: the EBLs in vesper bats were inserted approximately 14.2 to 53 million years ago. These results suggest that vesper bats have been repeatedly infected by bornaviruses at different points in time during evolution. This study provides novel insights into the evolutionary history of bornaviruses and demonstrates the robustness of combining database searches, PCR, and sequencing approaches to estimate insertion dates of bornaviruses.

KEY WORDS: endogenous bornavirus-like element, paleovirology, vesper bat

Viral sequences can be integrated into the genome of a host either as an essential replication step for viral multiplication or as an non-viral lifecycle dependent insertion, such as non-homologous end joining or retrotransposon-mediated insertions [9]. When such insertions occur within the host germline cells, the inserted sequences can be passed on to the offspring of the host through Mendelian inheritance [17, 29]. These integrated viral sequences are called endogenous viral elements (EVEs), and numerous EVEs have been identified in eukaryotic genomes [1, 9, 19]. Interestingly, some EVEs are known to be indispensable for various physiological functions in their hosts [16, 23]. Thus, integration of viral sequences into host genomes can be a driving force for the evolution of eukaryotes. Additionally, EVEs can serve as sources of information about ancient viruses [5, 7, 15, 26]. Unlike non-viral organisms, body fossils of viruses do not exist. Therefore, it is difficult to investigate ancient viruses. However, because many EVEs have been integrated more than millions of years ago, they serve as viral molecular fossils and can serve as good clues to understanding the evolutionary history of viruses [1].

Bornaviruses are non-segmented negative-strand RNA viruses, belonging to the order *Mononegavirales* [28]. Although they do not require reverse transcription and integration of their genomes into the host genome for replication, numerous bornavirus-derived EVEs, named endogenous bornavirus-like elements (EBLs), have been found in animal genomes [2, 12, 19]. Thus far, bornaviral N (nucleoprotein), M (matrix protein), G (glycoprotein), and L (RNA-dependent RNA polymerase) gene-derived EBLs have been reported and designated as EBLN, EBLM, EBLG, and EBLL, respectively [2, 5, 7, 10, 12–15, 19, 20]. Intriguingly, some EBLs seem to play functional roles in their hosts as RNAs or proteins [6, 13, 20, 24, 27], suggesting that bornaviruses have contributed to the evolution of their hosts. Additionally, as molecular fossils, EBLs have provided insights into ancient bornaviruses. EBLs have been found in diverse animal species, including those among which exogenous bornaviruses have not been detected [2, 5, 7, 10, 12–15, 19, 20]. This suggests that the historical host range of bornaviruses is much broader than

*Correspondence to: Horie, M.: horie.masayuki.3m@kyoto-u.ac.jp

©2018 The Japanese Society of Veterinary Science

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives (by-nc-nd) License. (CC-BY-NC-ND 4.0: https://creativecommons.org/licenses/by-nc-nd/4.0/)

that previously expected. Moreover, many EBLs were endogenized more than millions of years ago [2, 5, 7, 10, 12–14, 19, 20]. Surprisingly, an afrotherian EBLN was reported to have integrated more than 65 million years ago (MYA) [19, 20]. Therefore, the history of bornaviruses can be dated back to more than 65 MYA.

Bats are mammals belonging to the order Chiroptera, which comprises approximately 20% of classified mammalian species [25]. They serve as reservoirs of the highly lethal zoonotic mononegaviruses, such as rabies virus, Hendra virus, Nipah virus, Ebola virus, and Marburg virus [3, 8]. Although exogenous bornaviruses have not yet been detected in bats, many EBLs exist in bat genomes, especially in vesper bats (the family Vespertilionidae) such as bats of the genus *Eptesicus* and *Myotis* [5, 13, 14]. This suggests that bornaviruses infected ancestral vesper bats in the past. Additionally, molecular evolutionary analyses suggested that bornaviruses have been associated with bats for 70 MY [5]. However, insertion dates of the EBLs were not well analyzed: only one insertion date was estimated, and only sequences in the publicly available databases were used for that study [5]. Currently, the sequences of only 13 bat species are available in the NCBI whole genome shotgun (WGS) database (as of February 2018), although more than 1200 bat species have been described [25]. Thus, database searches are sometimes not enough to determine the ages of EBLs because they provide imprecise data ranges. We previously reported that an EBLL in bats of the genus *Eptesicus*, named eEBLL-1, was endogenized at least 11.8 MYA through a combination of PCR and sequencing approaches, using genomic DNA samples from *E. nilssonii* and *E. serotinus*; species whose genome sequences are not available in the public database [13]. Additionally, ages of EBLNs in snakes were also determined with a similar approach [7]. Thus, such approaches are powerful for the performance of paleovirological studies at high resolution.

Here, we comprehensively analyzed insertion dates of EBLs in vesper bats by gene orthology analyses, using sequence data in the public databases and those identified with PCR and sequencing. We successfully estimated the minimum ages of 10 EBLs and both the minimum and maximum ages of 3 EBLs in bats of the genus *Myotis* and *Eptesicus.* These findings provide novel insights into the paleovirology of bornaviruses as well as the co-evolution between bats and bornaviruses.

MATERIALS AND METHODS

tBLASTn search for the screening of EBLs in bats

A tBLASTn search was performed against the NCBI WGS database for chiropterans (taxid: 9397) on April 25, 2017, using the amino acid sequences of Borna disease virus 1 (BoDV-1) (accession number: AJ311522) as queries. In this study, sequence hits with an e-value threshold of 10^{-10} and 10^{-20} were defined as EBLNs and EBLLs, respectively. Consecutive sequence hits in contigs located in reasonable positions were regarded as fragmented EBL elements.

Preparation of genomic DNA samples

Genomic DNA of *E. nilssonii* was isolated from HAMOI-Enk cells, which were derived from the kidney [11], using QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany). Genomic DNA of *M. daubentonii* and *M. nattereri* were kindly provided by Dr. Marcel Alexander Müller.

PCR and direct sequencing

All PCR amplifications were performed with Phusion Green Hot Start II High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, U.S.A.) according to the manufacturer's instructions. Briefly, the PCR reactions were carried out using 30 *ng* of genomic DNA in a final volume of 20 μl, containing 1 U Phusion DNA polymerase, 0.2 mM dNTP, 0.5 nM gene-specific primers, and 3% volume DMSO under the following conditions: denaturing at 98°C for 30 sec, 35 cycles at 98°C for 10 sec, 70°C for 30 sec, 70°C for 30 sec per kilobase (kb), followed by an extended elongation at 72°C for 10 min. The PCR products were analyzed by agarose gel electrophoresis. The amplicons were purified using the innuPREP PCRpure kit (Analytik Jena AG, Jena, Germany), and then sequenced at Fasmac Co., Ltd. (Atsugi, Japan). The sequences of primers used for successful amplification of EBLs are listed in Table S1. The other primer sequences (negative for PCR or sequencing primers) are available upon request. The determined sequences were deposited in DDBJ (accession numbers LC378417-LC378424).

Gene orthology analyses

Gene orthologies were determined with a combination of EBLs nucleotide sequences alignment and their flanking sequences, and detection of repetitive elements in the flanking regions. First, repetitive elements were searched by RepeatMasker (Smit, AFA, Hubley, R & Green, P., http://www.repeatmasker.org). The nucleotide sequences and the annotation files (GFF files) were then imported to Geneious version 10.1.3 (Biomatters Ltd., Auckland, New Zealand; www.geneious.com). These sequences were aligned with the MAFFT [18] plugin in Geneious using the E-INS-i algorithm. Finally, gene orthologies were determined by analyzing conserved nucleotide sequences and conserved patterns of repetitive elements (RepeatMasker score >300) in the flanking regions.

EBL-empty loci (syntenic loci lacking EBL) were also determined by a combination of alignment and detection of repetitive elements as described above. BLASTn searches were performed against the WGS and RefSeq genome databases for Chiroptera (taxid: 9397) using EBLs and their flanking sequences as queries. The BLAST hits were retrieved and analyzed by RepeatMasker and multiple alignments as described above. Genomic loci lacking corresponding EBLs, but sharing similarities with flanking regions of EBLs were regarded as EBL-empty loci. All the multiple alignments are available upon request.

Table 1. Summary of the gene orthology analyses for EBLN

Name of EBL	*Myotis*				*Eptiscus*	
	myEBLN-1	myEBLN-2	myEBLN-3	myEBLN-4	eEBLN-1	eEBLN-2
Myotis davidii				+		Empty
Myotis brandtii	+	+	+		Empty	
Myotis lucifugus	+	+				
Myotis daubentonii	+		+			
Myotis nattereri	+		+			
Eptesicus fuscus	Empty	Empty			+	+
Eptesicus nillsonii					+	
Estimated age (MYA)	16.2–31	14.2–31	At least 16.2	Not determined	15.3–31	Not determined

+, positive for an orthologous EBL; Empty, presence of an empty locus (a sytenic locus without EBL).

Estimation of insertion time of EBLs

Insertion dates of EBLs were determined based on the gene orthologies. The divergent times of host species were estimated from TimeTree [21].

Phylogenetic analyses

Nucleotide sequences of EBLNs or EBLLs in vesper bats were aligned by MAFFT using the E-INS-i algorithm [18]. Phylogenetic trees were inferred by the neighbor-joining method with p-distance using MEGA7 [22] The reliability of each internal branch was evaluated by 1,000 bootstrap replicates.

RESULTS

Detection of EBLs in genomes of vesper bats in the WGS database

For screening of EBLs in vesper bats, we performed a tBLASTn search using the amino acid sequence of each gene in BoDV-1. We found two EBLNs and three EBLLs in *E. fuscus*, and six EBLNs and 16 EBLLs in three species of bats of the genus *Myotis* (Tables S2 and S3). Except for an EBLL (myEBLL-4) in the genome of *M. brandtii* (Table S3), all the BLAST hits had already been identified as EBLs in previous studies [5, 13, 14].

Gene orthology analyses

Next, we tried to determine the ages of the detected EBLs. Whereas minimum ages of EBLs can be estimated by searching for orthologues, maximum ages of EBLs can be determined by identifying EBL-empty loci (that is, syntenic loci without corresponding EBLs) [1, 7]. Therefore, we determined the gene orthologies of the detected EBLs and searched for EBL-empty loci through a combination of database searches, PCR, and sequencing analyses. We performed BLASTn against all available bat WGS sequences (the order Chiroptera) in the NCBI database, using the sequences of the EBLNs and their flanking regions as queries. We retrieved the sequence hits and analyzed the sequence similarities and patterns of repetitive elements in the sequences (Tables 1 and 2). Using these data, we estimated the ages of EBLs in vesper bats as described chronologically below in detail.

Inter-family and inter-genus EBLs in bat genomes

We found EBLLs that were shared across several families of bats. EBLLs detected in the three species of *Myotis* by tBLASTn (in ANKR01171284, ALWT01098736, and AAPE02024702) seemed to be orthologous (Fig. S1a). This EBLL is also shared by *E. fuscus*, *Miniopterus natalensis*, and *Pteronotus parnellii* (Fig. S1a). Thus, this EBLL may have been endogenized prior to the divergence of *Myotis*, *Eptesicus*, *Miniopterus*, and *Pteronotus*, which is estimated to be 53 MYA (Fig. 1 and Table 2). This estimation is identical to a previous report that suggested this EBLL became endogenous at least 50 MYA, although that report described that this EBLL exists only in *M. davidii* and *P. parnellii* [5].

Next, we found that another EBLL is orhologous between the genera *Myotis* and *Eptesicus*. This inter-genus EBLL is shared by *E. fuscus* and *M. davidii* (Fig. S1b). We could not find any empty loci for this EBLL. Thus, this EBLL became endogenous prior to the divergence of bats of the genera *Myotis* and *Eptesicus*, which is estimated to be 31 MYA (Fig. 1 and Table 2).

Genus-specific EBLs in the genomes of Myotis bats

Three lineages of EBLLs (named myEBLL-2, 3 and 4) are shared in three species of bats from the genus *Myotis* (Fig. S2a–c), suggesting that these EBLLs have endogenized before the divergence of these three species. Although some contigs in *E. fuscus* may be empty loci for myEBLL-2, 3 and 4, we could not determine any empty loci because of the short lengths of the contigs (data not shown). Thus, we concluded that myEBLL-2, 3 and 4 may have become endogenous at least 20.8 MYA (Fig. 1 and Table 2).

We found that myEBLL-5 is orthologous between *M. davidii*, and *M. nattereri* whereby we detected myEBLL-5 in *M. davidii* by tBLASTn, and detected its orthologue in *M. nattereri* with PCR and sequencing (Figs. S1d and S5d). Thus, myEBLL-5 was integrated into the bat genome prior to the divergence of these two species, which is estimated to be 20.8 MYA (Fig. 1 and

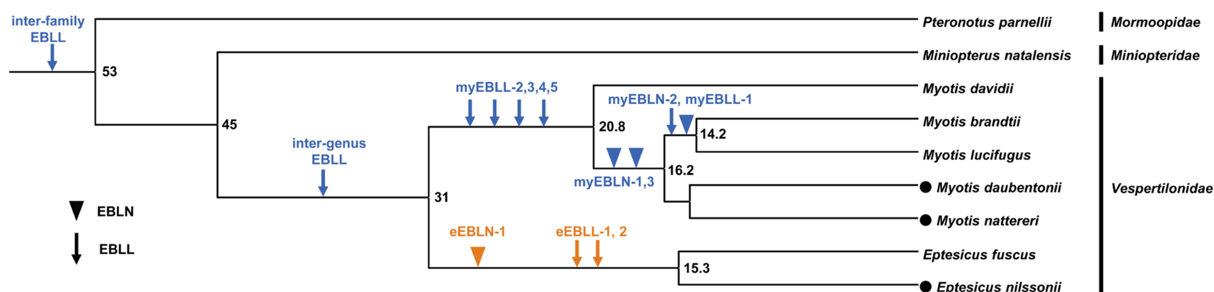

Fig. 1. Bat evolution and integration of bornavirus-derived genes. A schematic phylogenetic tree of bats and endogenization events of bornaviral sequences are shown. Arrowheads and arrows indicate estimated endogenization events of EBLNs and EBLLs, respectively. Blue arrowheads and arrows show EBL insertions for which we could estimate only minimum ages. Orange ones indicate EBL integrations for which we could estimate both maximum and minimum ages. Species whose genomic DNA were used for the PCR and sequencing analyses are marked with circles. Divergent times are indicated in each internal node. The topology of the phylogenetic tree and the divergent times of bats were taken from TimeTree. Note that the insertion dates of some EBLs may have been underestimated (see Discussion).

Table 2. Summary of the gene orthology analyses for EBLL

Name of EBL	Inter-family	Inter-genus	*Myotis*						*Eptesicus*	
	Undefined	Undefined	myEBLL-1	myEBLL-2	myEBLL-3	myEBLL-4	myEBLL-5	myEBLL-6	eEBLL-1	eEBLL-2
Myotis davidii	+	+		+	+	+	+	+		Empty
Myotis brandtii	+		+	+	+	+			Empty	
Myotis lucifugus	+		+	+	+	+				
Myotis daubentonii										
Myotis nattereri							+			
Eptesicus fuscus	+	+	Empty						+[a)]	+
Eptesicus nillsonii									+[a)]	+
Miniopterus natalensis	+									
Pteronotus parnellii	+									
Estimated age (MYA)	At least 53	At least 31	14.2–31	At least 20.8	At least 20.8	At least 20.8	At least 20.8	Not determined	15.3–31	15.3–31

+, positive for an orthologous EBL; Empty, presence of an empty locus (a sytenic locus without EBL). a) These EBLs were anlayzed in the previous study [13].

Table 2). However, myEBLL-5 was not detected in *M. brandtii* and *M. lucifugus*, although it was excepted to be present based on the assumptions that endogenization occurred in the common ancestor (Fig. 1).

Genus-specific EBLs in the genomes of Eptesicus bats

We identified that eEBLN-1 and eEBLL-2 are shared between *E. fuscus* and *E. nilssonii* by PCR and sequencing (Figs. S2e, S2g and S6). Additionally, through database searches, we found empty loci for eEBLN-1 and eEBLL-2 in *M. brandtii* and *M. davidii*, respectively (Fig. S2e and S2g).

We previously found that eEBLL-1 had been endogenized prior to the divergence of *E. fuscus* and *E. nilssonii* [13]. Here, we further identified an empty locus in a contig of *M. brandtii* which highly likely lacked EBLLs orthologous to eEBLL-1 (Fig. S2f). Although nucleotide sequences upstream of the target site duplication (TSD) of eEBLL-1, which can be the hallmark of the connection between host- and virus-derived sequences, are highly conserved between *E. fuscus* and *M. brandtii*, those downstream of the TSD are not alignable (Fig. S2f). Although we could not analyze the nucleotide sequence downstream of eEBLL-1 because of the limited length of the contig from *E. fuscus*, these observations suggest that eEBLL-1 was inserted into this region.

Taken together, eEBLN-1, eEBLL-1, and eEBLL-2 seems to have become endogenous before the divergence of *E. fuscus* and *E. nilssonii* and after the divergence of the genera *Eptesicus* and *Myotis*, indicating that the ages of these EBLs are estimated to be 15.3–31 MY (Fig. 1, Tables 1 and 2).

Sub-genus specific EBLs in bats of the genus Myotis

We found that myEBLN-1is shared by four species of *Myotis* bats (*M. brandtii*, *M. lucifugus*, *M. daubentonii*, and *M. nattereri*). Through database searches, we revealed that myEBLN-1 is orthologous between *M. brandtii* and *M. lucifugus* (Fig. 2c). We also performed PCR analyses for these EBLNs using genomic DNA of *M. daubentonii* and *M. nattereri*, and detected bands of expected sizes whose sequences were almost identical to myEBLN-1 (Fig. 2a–c). Additionally, we found empty loci for myEBLN-1 in the genome of *E. fuscus* (Fig. 2c). These results suggest that myEBLN-1 was endogenized prior to the divergence of the four species

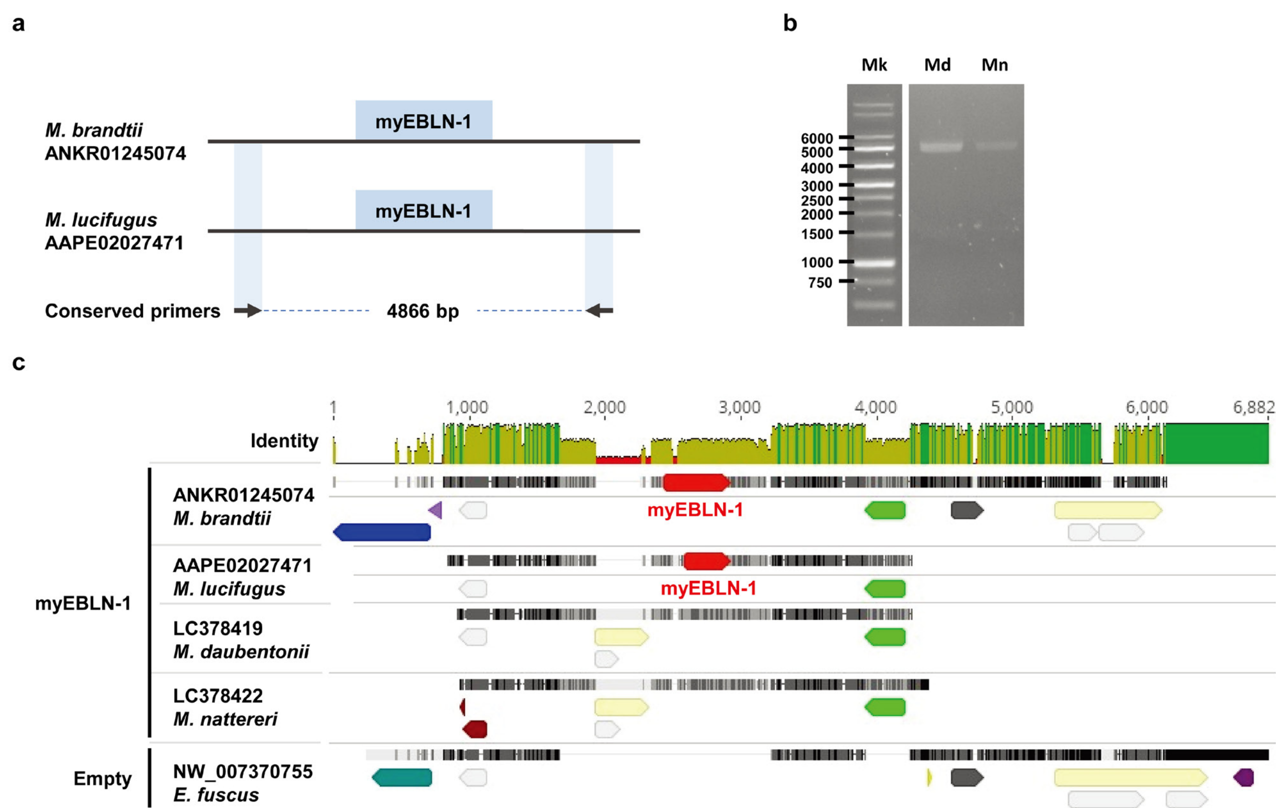

Fig. 2. Representative gene orthology analysis of an EBL. (a) A schematic diagram of the PCR analysis used to detect the myEBLN-1 and its flanking region. Primer pairs used for the PCR are depicted as arrows. The expected size (base pairs, bp) of the amplicon is shown. (b) PCR amplification of myEBLN-1 loci from genomic DNA of *M. daubentonii* and *M. nattereri*. Mk, DNA ladder marker; Md, *M. daubentonii*; Mn, *M. nattereri*. Note that the pictures were cropped from the original picture. The original picture is available upon request. (c) An alignment of myEBLN-1 and its empty loci. The names of species and the accession numbers or NCBI RefSeq numbers are shown on the left side. Sequence similarity is represented by the density of black color, and thin horizontal lines show gaps. EBLs detected by the tBLASTn screening are depicted as red arrow blocks, and repetitive elements are represented as arrow blocks below each of the sequences. The colors of the arrow blocks indicate the types of repetitive elements.

and after the divergence of *Myotis* and *Eptesicus*, which are estimated to be 16.2 and 31 MYA, respectively (Fig. 1 and Table 1).

We found that myEBLN-3 is shared by *M. brandtii*, *M. dauubentonii*, and *M. nattereri* as follows. Although we could not identify any orthologue for myEBLN-3 using database searches, we detected myEBLN-3 orthologues in the genomes of *M. daubentonii* and *M. nattereri* using PCR and sequencing (Figs. S3a and S5b). Thus, the age of myEBLN-3 is more than 16.2 MY (Fig. 1 and Table 1). However, we could not detect an orthologue of myEBLN-3 in the genome of *M. lucifugus,* although it was theoretically expected to be present (see Discussion).

Through database searches, we found that myEBLN-2 and myEBLL-1 are orthologous between *M. brandtii* and *M. lucifugus*, and their empty loci are present in the genome of *E. fuscus* (Fig. S3b and S3c). Our data suggests that these EBLs were integrated before the divergence of *M. brandtii* and *M. lucifugus*, and after the divergence of the genus *Myotis* and *Eptesicus*, which are estimated to be 14.2 and 31 MYA, respectively (Fig. 1, Tables 1 and 2). We note that although we could not detect any band by PCR using genomic DNA of *M. daubentonii* and *M. nattereri* with primers binding to the flanking regions, we detected the expected bands when we used the same genomic DNA and performed the PCR with primers designed for myEBLN-2 (Fig. S5a and Table S4). The sequences of the bands showed a 94.7% identity to myEBLN-2. This may have been caused by a genomic deletion of the flanking region or may be due to independent integrations of similar viruses. However, because of their lack of flanking sequences, we could not further assess the above possibilities.

EBLs lacking orthologues

For myEBLN-4, myEBLL-6, and eEBLN-2, we could not find any orthologues either from database searches or PCR. Regarding myEBLN-4 in *M. davidii*, we found very similar EBLNs in the genomes of *M. brandtii*, *M. lucifugus*, *M. daubentonii*, and *M. nattereri* with database searches and PCR (Figs. S4a, S5c and Table S4). However, we could not determine any gene orthology because the contigs and determined sequences were short, and thus, we could not analyze their flanking regions. myEBLL-6 was found in the genome of *M. davidii* by tBLASTn screening, but its orthologue was not detected in other bats in the genus *Myotis*

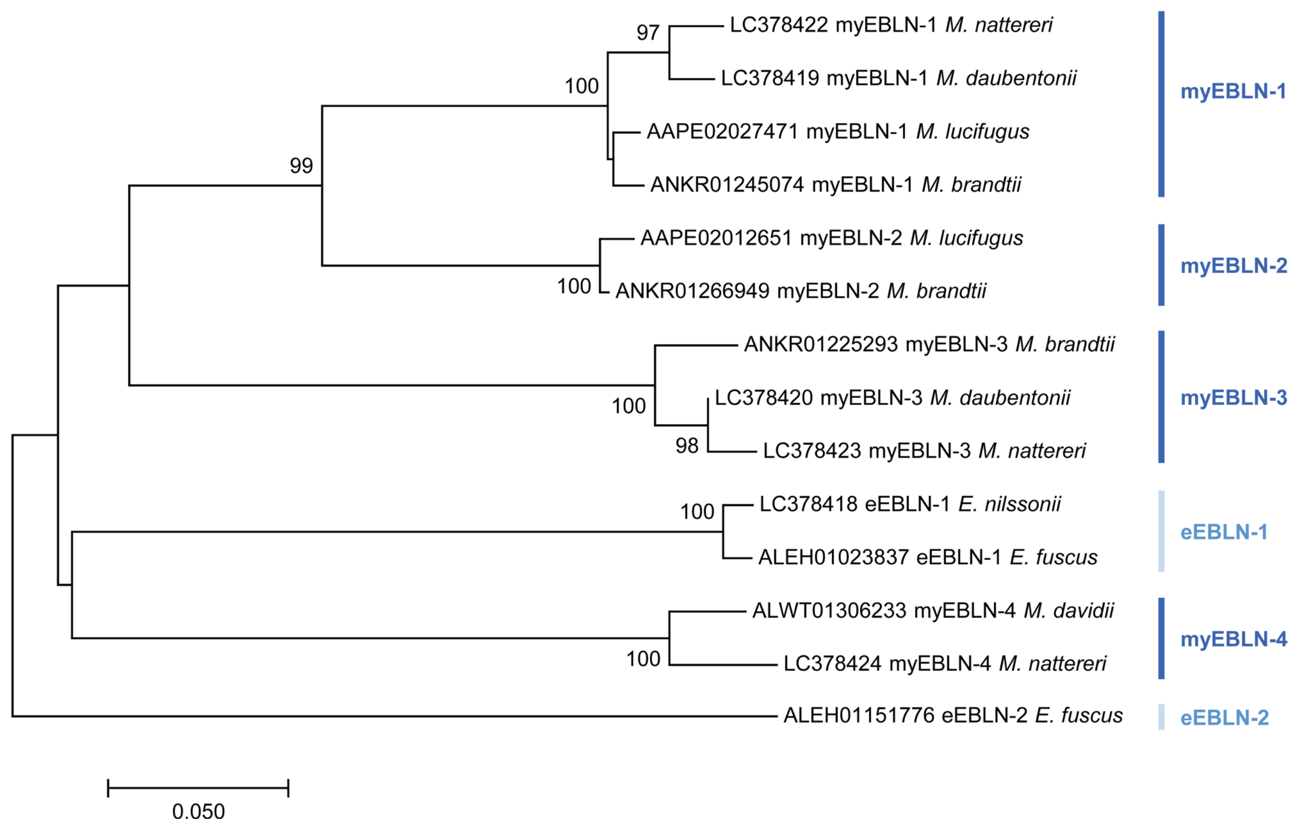

Fig. 3. Phylogenetic tree of EBLNs in vesper bats. The phylogenetic tree was constructed based on nucleotide sequences of EBLNs in vesper bats using the neighbor-joining method with 1,000 bootstrap replicates. Bootstrap values more than 90% are indicated. Scale bar shows nucleotide substitutions per site.

either by database search or PCR. Intriguingly, a nucleotide sequence that is similar to the myEBLL-6 sequence (nucleotide identity: 89.1%) was detected in *E. fuscus* (Fig. S4b and Table S5). Although sequences flanking to the BLAST hit region were also very similar, we could not determine the gene orthology because we could not exclude the possibility that these regions were derived from bornavirus L gene but not detected as EBLL by tBLASTn due to low sequence similarity. In this case, the BLAST hit region corresponds to 770–922 amino acid residues of BoDV-1 L protein that consists of 1710 amino acids, and the flanking regions do not share any repetitive elements. Thus, more than 2000 nucleotides of upstream and downstream of the BLAST hit region might be derived from bornavirus L gene. As for eEBLN-2, although we found an empty locus in *M. davidii*, we could not detect any orthologue of eEBLN-2 in other species (Fig. S4c and Table S4).

Phylogenetic analyses

To evaluate the phylogenetic relationships of EBLs, we performed phylogenetic analyses. Phylogenetic trees showed that each lineage of the EBLs forms a robust monophyletic clade (Figs. 3 and S7). The EBLN in *M. nattereri* that is very similar to myEBLN-4 in *M. davidii* but for which gene orthology could not be determined (see the subsection *EBLs lacking orthologues*), is clustered phylogenetically with myEBLN-4.

DISCUSSION

In this study, we systematically investigated the ages of EBLs in vesper bats. We successfully determined the orthologies of many EBLs, which enabled us to investigate ancient bornaviruses at a higher resolution than in previous studies. Although several studies reported the presence of EBLs in vesper bats, only the ages of eEBLL-1 and an inter-family EBLL have been determined [5, 13]. One study, investigating the evolution of bat EBLs, suggested the long co-evolutionary history between bats and bornaviruses [5]. However, there was a limitation in that study: only sequences in the public database were used for the analyses, and thus, the insertion dates of EBLs were not precisely determined. To overcome this limitation, we adopted a combination of database searches, PCR, and a sequencing strategy, using the genomic DNA of bats. This allowed us to estimate the insertion dates of 13 EBLs in vesper bats. The integration events seemed to have occurred at different points during evolution, suggesting that vesper bats may have been repeatedly infected by bornaviruses over a long period of time. Thus, this study contributes to a deep understanding of the paleovirology of bornaviruses.

Here, we estimated the ages of EBLs in bats of the genus *Myotis* and *Eptesicus*; however, there is a limitation to this study: we may have underestimated the ages of some of the EBLs because of the following reasons. First, we could not determine gene orthologies for some EBLs because some contigs were too short to analyze their flanking sequences. In addition, for some EBLs, we could not amplify the DNA samples with the primers binding to the flanking regions, but could do so with primers binding to the EBL regions (e.g. myEBLN-2; Table S4 and Fig. S3b). This might be due to mismatches between our primers used and the flanking sequences. Although we also tested another primer pair binding to each of the flanking regions of these EBLs, we could not obtain any bands following gel electrophoresis (data not shown). Thus, the sequences of these flanking regions might be different from their corresponding sequences in other species due to mutations, insertions, deletions and/or recombination. Second, we realized that some species seemed to lack EBLs that theoretically should have existed in the genome. For example, we could not detect any myEBLN-3 orthologue in *M. lucifugus*, although this EBLN seems to have been integrated before the divergence of *M. brandtii*, *M. lucifugus*, *M. daubentonii* and *M. nattereri* (Table 1, Figs. 1 and S3a). These observations could be attributed to genomic deletions after endogenization or incomplete lineage sorting, which cause discordance of species trees and gene trees as also proposed for endogenous retroviruses [30]. Alternatively, the WGS sequences might be incomplete. Further improvement in the quality of genome sequences will contribute to understanding paleovirology in more detail.

Interestingly, most of the EBLs seemed to have been integrated into the bat genomes after the divergence of their genera (Fig. 1). This may be due to the activities of retrotransposons. Because bornaviruses do not encode either reverse-transcriptase or integrase, endogenization of bornavirus-derived sequences was considered to have been mediated by reverse-transcriptase of retrotransposons in the genomes of the hosts [2, 12]. Indeed, it was reported that some anthropoid EBLNs became endogenous when LINE-1 was active in the hosts [9]. In addition, we and others previously reported that poly-A stretches and target site duplications, which are known to be footprints of LINE-1-mediated integrations, are located near several EBLs. Therefore, it is plausible that LINE-1 was highly active when EBLs became endogenous in vesper bats. Alternatively, bats may have been frequently infected by bornaviruses during that era. Further analyses, such as the investigation of the ancient LINE-1 activity in vesper bats, like Cantrell *et al*. have done in megabats are needed to assess these hypotheses [4].

Here, we identified EBL orthologues in several species whose genome sequences are not available in the public database. Availability of sequences of EBL orthologues contributes to investigations of proteins encoded by ancient bornaviruses. Westerman *et al*. reconstructed the sequence of an ancient parvovirus from parvoviral EVEs and analyzed the property of the protein. Such analyses would provide information about ancient viral proteins [26]. Accumulation of such data for diverse EVEs will be useful to understanding paleovirology. Moreover, identification of EBL orthologues can be used for the examination of natural selection in EBLs. As previously reported, several EBLs were reported to have evolved under negative selection [13, 20]. Thus, the identified sequence information can serve as clues for understanding the biological significance of EBLs in their host species.

This study also provides insights into the co-evolution between bornaviruses and vesper bats. As described previously, eEBLL-1, which is referred to as efEBLL-1 in this study and is highly conserved among bats of the genus *Eptesicus*, retains a large and intact ORF derived from an RNA-dependent RNA polymerase gene of an ancient bornavirus [13]. Interestingly, among the two EBLNs and four EBLLs in bats of the genus *Eptesicus*, only eEBLL-1 has maintained the intact ORF, whereas those of the other EBLs were disrupted during the evolution. Again, this strongly suggests that eEBLL-1 encodes a functional protein in bats of the genus *Eptesicus*; regarding the other EBLs, they may have become pseudogenes. Alternatively, they may encode non-coding RNAs or truncated proteins. Further studies, such as transcriptome and biological analyses, are necessary to understand the functions of EBLs in these bats.

Taken together, this study has expanded our knowledge of ancient bornavirus infections during evolution as well as the co-evolution between bats and bornaviruses. Similar studies would have to be conducted with other non-retroviral EVEs to contribute to the understanding of the evolutionary history of bornaviruses.

ACKNOWLEDGMENTS. We thank Dr. Keiko Takemoto for technical assistance. We are also grateful to Dr. Marcel Alexander Müller for providing the genomic DNA samples. This study is supported by the Hakubi project (MH), 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 JP 17am0301015h0003) (KT).

REFERENCES

1. Aiewsakun, P. and Katzourakis, A. 2015. Endogenous viruses: Connecting recent and ancient viral evolution. *Virology* **479–480**: 26–37. [Medline] [CrossRef]
2. Belyi, V. A., Levine, A. J. and Skalka, A. M. 2010. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS Pathog.* **6**: e1001030. [Medline] [CrossRef]
3. Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V. and Schountz, T. 2006. Bats: important reservoir hosts of emerging viruses. *Clin. Microbiol. Rev.* **19**: 531–545. [Medline] [CrossRef]
4. Cantrell, M. A., Scott, L., Brown, C. J., Martinez, A. R. and Wichman, H. A. 2008. Loss of LINE-1 activity in the megabats. *Genetics* **178**: 393–404. [Medline] [CrossRef]
5. Cui, J. and Wang, L. F. 2015. Genomic Mining Reveals Deep Evolutionary Relationships between Bornaviruses and Bats. *Viruses* **7**: 5792–5800. [Medline] [CrossRef]

6. Fujino, K., Horie, M., Honda, T., Merriman, D. K. and 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. [Medline] [CrossRef]
7. Gilbert, C., Meik, J. M., Dashevsky, D., Card, D. C., Castoe, T. A. and Schaack, S. 2014. Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc. Biol. Sci.* **281**: 20141122. [Medline] [CrossRef]
8. Han, H. J., Wen, H. L., Zhou, C. M., Chen, F. F., Luo, L. M., Liu, J. W. and Yu, X. J. 2015. Bats as reservoirs of severe emerging infectious diseases. *Virus Res.* **205**: 1–6. [Medline] [CrossRef]
9. Horie, M. and Tomonaga, K. 2011. Non-retroviral fossils in vertebrate genomes. *Viruses* **3**: 1836–1848. [Medline] [CrossRef]
10. Horie, M. and Tomonaga, K. 2018. Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses. *Virus Res.* [Medline] [CrossRef]
11. Horie, M., Akasaka, T., Matsuda, S., Ogawa, H. and Imai, K. 2016. Establishment and characterization of a cell line derived from Eptesicus nilssonii. *J. Vet. Med. Sci.* **78**: 1727–1729. [Medline] [CrossRef]
12. Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., Ikuta, K., Jern, P., Gojobori, T., Coffin, J. M. and Tomonaga, K. 2010. Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* **463**: 84–87. [Medline] [CrossRef]
13. Horie, M., Kobayashi, Y., Honda, T., Fujino, K., Akasaka, T., Kohl, C., Wibbelt, G., Mühldorfer, K., Kurth, A., Müller, M. A., Corman, V. M., Gillich, N., Suzuki, Y., Schwemmle, M. and Tomonaga, K. 2016. An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. *Sci. Rep.* **6**: 25873. [Medline] [CrossRef]
14. Horie, M., Kobayashi, Y., Suzuki, Y. and Tomonaga, K. 2013. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**: 20120499. [Medline] [CrossRef]
15. Hyndman, T. H., Shilton, C. M., Stenglein, M. D. and Wellehan, J. F. X. Jr. 2018. Divergent bornaviruses from Australian carpet pythons with neurological disease date the origin of extant Bornaviridae prior to the end-Cretaceous extinction. *PLoS Pathog.* **14**: e1006881. [Medline] [CrossRef]
16. Imakawa, K., Nakagawa, S. and Miyazawa, T. 2015. Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placentation during mammalian evolution. *Genes Cells* **20**: 771–788. [Medline] [CrossRef]
17. Jern, P. and Coffin, J. M. 2008. Effects of retroviruses on host genome function. *Annu. Rev. Genet.* **42**: 709–732. [Medline] [CrossRef]
18. Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**: 772–780. [Medline] [CrossRef]
19. Katzourakis, A. and Gifford, R. J. 2010. Endogenous viral elements in animal genomes. *PLoS Genet.* **6**: e1001191. [Medline] [CrossRef]
20. Kobayashi, Y., Horie, M., Nakano, A., Murata, K., Itou, T. and Suzuki, Y. 2016. Exaptation of bornavirus-like nucleoprotein elements in afrotherians. *PLoS Pathog.* **12**: e1005785. [Medline] [CrossRef]
21. Kumar, S., Stecher, G., Suleski, M. and Hedges, S. B. 2017. Timetree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**: 1812–1819. [Medline] [CrossRef]
22. Kumar, S., Stecher, G. and Tamura, K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**: 1870–1874. [Medline] [CrossRef]
23. Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., LaVallie, E., Tang, X. Y., Edouard, P., Howes, S., Keith, J. C. Jr. and McCoy, J. M. 2000. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **403**: 785–789. [Medline] [CrossRef]
24. Myers, K. N., Barone, G., Ganesh, A., Staples, C. J., Howard, A. E., Beveridge, R. D., Maslen, S., Skehel, J. M. and Collis, S. J. 2016. The bornavirus-derived human protein EBLN1 promotes efficient cell cycle transit, microtubule organisation and genome stability. *Sci. Rep.* **6**: 35548. [Medline] [CrossRef]
25. Simmons, N. B. 2005. Order chiroptera. pp. 312–529 *In*: Mammal Species of the World. A Taxonomic and Geographic Reference (Wilson, D. E. and Reeder, D. M. E. eds.), Johns Hopkins University Press, Washington, D.C.
26. 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., Agbandje-McKenna, M., Kotin, R. M. and Alexander, I. E. 2016. Germline viral "fossils" guide in silico reconstruction of a mid-Cenozoic era marsupial adeno-associated virus. *Sci. Rep.* **6**: 28965. [Medline] [CrossRef]
27. Sofuku, K., Parrish, N. F., Honda, T. and Tomonaga, K. 2015. Transcription Profiling Demonstrates Epigenetic Control of Non-retroviral RNA Virus-Derived Elements in the Human Genome. *Cell Reports* **12**: 1548–1554. [Medline] [CrossRef]
28. Tomonaga, K., Kobayashi, T. and Ikuta, K. 2002. Molecular and cellular biology of Borna disease virus infection. *Microbes Infect.* **4**: 491–500. [Medline] [CrossRef]
29. Weiss, R. A. 2006. The discovery of endogenous retroviruses. *Retrovirology* **3**: 67. [Medline] [CrossRef]
30. Zhuo, X. and Feschotte, C. 2015. Cross-species transmission and differential fate of an endogenous retrovirus in three mammal lineages. *PLoS Pathog.* **11**: e1005279. [Medline] [CrossRef]