# The hidden diversity of ancient bornaviral sequences from X and P genes in vertebrate genomes

Bea Clarise B. Garcia [1], Yahiro Mukai [2,3], Keizo Tomonaga [2,3,4], & Masayuki Horie [1,5*]

[1] Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, Izumisano, Osaka, Japan

[2] Laboratory of RNA Viruses, Institute for Frontier Life and Medical Sciences, Kyoto University, Kyoto, Japan

[3] Laboratory of RNA Viruses, Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University

[4] Department of Molecular Virology, Graduate School of Medicine, Kyoto University

[5] Osaka International Research Center for Infectious Diseases, Osaka Metropolitan University, Izumisano, Osaka, Japan

***Corresponding author**

Email: mhorie@omu.ac.jp

**Abstract**

Endogenous bornavirus-like elements (EBLs) are heritable sequences derived from bornaviruses in vertebrate genomes that originate from transcripts of ancient bornaviruses. EBLs have been detected using sequence similarity searches such as tBLASTn, whose technical limitations may hinder the detection of EBLs derived from small and/or rapidly evolving viral X and P genes. Indeed, no EBLs derived from the X and P genes of orthobornaviruses have been detected to date in vertebrate genomes. Here, we aimed to develop a novel strategy to detect such "hidden" EBLs. To this aim, we focused on the 1.9 kb read-through transcript of orthobornaviruses, which encodes a well-conserved N gene and small and rapidly evolving X and P genes. We show a series of evidence supporting the existence of EBLs derived from orthobornaviral X and P genes (EBLX/Ps) in mammalian genomes. Furthermore, we found that an EBLX/P is expressed as a fusion transcript with the cellular gene, *ZNF451*, which potentially encodes the ZNF451/EBLP fusion protein in miniopterid bat cells. This study contributes to a deeper understanding of ancient bornaviruses and co-evolution between bornaviruses and their hosts. Furthermore, our data suggest that endogenous viral elements detected thus far are just the tip of the iceberg, and further studies are required to understand ancient viruses more accurately.

## Introduction

Ancient viral sequences can be found within the genomes of organisms in the form of endogenous viral elements (EVEs) (1). EVEs are derived from the integration of viral sequences into the host germline DNA and transfer of the integrated sequences to subsequent progenies (1, 2). As molecular fossils of ancient viruses, EVEs have expanded our knowledge of the diversity, host range, and long-term evolution of viruses (1). Furthermore, some EVEs are co-opted by hosts to function as novel gene regulatory elements, non-coding RNAs, and proteins (2).

Most EVEs are derived from retroviruses, as the viruses reverse-transcribe and integrate the viral genomes into the host genomes using virally encoded enzymes. However, non-retroviral EVEs have also been identified. Endogenous bornavirus-like elements (EBLs) are EVEs derived from bornaviruses (family *Bornaviridae*), a group of non-segmented negative-strand RNA viruses (3, 4). Although bornaviruses are RNA viruses, evidence strongly suggests that bornaviral mRNA is reverse transcribed and integrated into the host genome via the activity of the long interspersed nuclear element 1 (LINE-1) retrotransposon (3, 5). We previously revealed an abundance of EBLs in vertebrate genomes derived from the bornaviral genes N, P, M, G, and L (S1 Fig), which are referred to as EBLN, EBLP, EBLM, EBLG, and EBLL, respectively (6). These EBLs originated from ancient bornaviruses of all genera in the family *Bornaviridae*: *Orthobornavirus*, *Carbovirus*, and *Cultervirus*.

Despite the rigorous search, there has been a bias in the detection of EBLs. EBLNs comprise the majority of the detected EBLs, while no EBLs derived from the X and P genes have been reported except for a few carboviral EBLPs (3, 6-10). We proposed two hypotheses to explain the lack of EBLX and EBLP. First, X/P mRNA, which

encodes the X and P proteins, may have a low propensity for reverse-transcription and/or integration into the host genome. However, previous studies of Borna disease virus 1 (BoDV-1; genus *Orthobornavirus*) showed that X/P mRNA is the most abundant transcript among the viral mRNAs (11, 12), which should therefore provide enough chances for integration. Moreover, we have previously demonstrated that the X/P mRNA of BoDV-1 can be reverse-transcribed and integrated into the genome of host cells (3, 7). This led us to the second hypothesis, which states that EBLs derived from the X and P genes are present but are difficult to detect due to technical limitations of the sequence similarity search. EBLs have been routinely searched using tBLASTn; however, this can be an unsuitable method for detecting small and rapidly evolving genes (8). Indeed, the X and P genes are small and the least conserved among bornaviral genes (Fig 1). If the second hypothesis is correct, some EBLs may remain undetected. Therefore, we could be missing important insights into ancient bornaviruses and the co-evolution of bornaviruses and their hosts.

In this study, we sought to develop a novel strategy to detect these hidden EBLs by focusing on the nature of the read-through transcription of orthobornaviruses that produce a transcript encoding a highly conserved N gene and rapidly evolving X and P genes. We show evidence supporting the existence of several EBLX/Ps in vertebrate genomes. Furthermore, an EBLX/P in miniopterid bats appear to be transcribed as chimeric transcripts with a host gene, suggesting a potential co-option of EBLs by the host.

## Results

### *Miniopterid bat EBLN is possibly derived from an N/X/P transcript*

We previously identified miEBLN-1, an EBLN derived from an ancient orthobornavirus that retains an intact N ORF in miniopterid bats (13). The miEBLN-1 locus is present in the genomes of miniopterid bats, specifically in those of *Miniopterus natalensis*, *M. schreibersii*, and *M. fuliginosus*, but is empty (i.e., syntenic loci lacking miEBLN-1) in the genomes of *Rhinolophus ferrumequinum* and *Phyllostomus hastatus*. We re-examined the presence of miEBLN-1 using the current genome datasets and found the miEBLN-1-empty locus in *Tadarida brasiliensis* (Fig 2a and S2a Fig).

Based on the nucleotide alignment, we deduced the insertion size of the miEBLN-1 locus to be 1.9 kb, consisting of a 1.2-kb miEBLN-1 region and a 0.7-kb unannotated region. Therefore, we carried out a series of analyses to understand the source of the 0.7-kb unannotated region. First, we performed a BLASTn search (14) against the NCBI nt database (15) using the nucleotide sequence of the 0.7-kb unannotated region as a query. The hits consisted only of miEBLN-1 sequences (data not shown). Next, we assessed whether the 0.7-kb unannotated region is derived from a short interspersed nuclear element (SINE) due to the precedence of chimeric integration comprising nucleotide sequences from an N transcript and SINE (3, 5). We performed a BLASTn search against the miniopterid bat genomes using the nucleotide sequence of the 0.7-kb region as a query. We only detected five BLAST hits in the genomes of *M. natalensis* and *M. schreibersii*, which we identified as EBLs in the later sections. We also analyzed the sequence of the 0.7-kb region by CENSOR (16), but no hits were detected. These results indicated that the 0.7-kb fragment is not derived from a SINE.

We noticed that the insertion, including miEBLN-1, is almost similar in size to the

N/X/P mRNA of extant orthobornaviruses (Fig 2a and S2a Fig), which led us to speculate that the insertion may be derived from an N/X/P transcript. N/X/P mRNA is produced by read-through of the termination signal T1, resulting in a 1.9 kb mRNA containing the transcription signal sequences S1, T1/S2, and T2 (Fig 2b) (12). Therefore, we investigated the presence of these signal-like sequences in the 1.9-kb miEBLN-1 insertion. We found transcription signal-like sequences at reasonable positions with respect to extant N/X/P transcripts (Fig 2c and S3b Fig). A poly-A stretch was also observed immediately downstream of the T2 signal, indicating that the insertion originated from an mRNA transcript (Fig 2c and S3b Fig). Furthermore, a tandem duplication sequence, TTGTGCTA/TTCTTGTA, was observed immediately upstream and downstream of S1 and T2-poly-A sequences, respectively (Fig 2c and S3b Fig). This is a potential target site duplication (TSD), which is a hallmark of LINE-1-mediated mRNA integration (17). These results suggested that the 1.9-kb insertion, including miEBLN-1, is derived from an ancient N/X/P transcript integrated by LINE-1. Therefore, we tentatively refer to this locus as miEBLN/X/P-1.

***Additional EBLs possibly derived from N/X/P transcripts in miniopterid bat genomes***

As described above, orthobornaviruses express mRNAs encoding N (1.2 kb), X and P (X/P mRNA, 0.7 kb), or N, X, and P (N/X/P mRNA, 1.9 kb) (Fig 2a). If miEBLN/X/P-1 is truly derived from N/X/P mRNA, there may be additional EBLs derived from N/X/P or X/P mRNA. To examine this possibility, we conducted a BLASTn search against bat genomes using the nucleotide sequence of miEBLN/X/P-1 as a query. We obtained six hits from the genomes of *M. natalensis* and *M. schreibersii* (Fig 3a and S1 Table). While four hits were aligned in the entire N/X/P region of miEBLN/X/P-1 (Fig 3a, red lines),

interestingly, two other hits were aligned only to the putative X/P region (Fig 3a blue lines). Note that the N regions of four hits aligned to entire miEBLN/X/P-1 were detected as EBLNs in a previous study (6).

We first assessed hits that aligned to the entire region of miEBLN/X/P-1. Four hits showed 1.9 kb alignment lengths and high sequence similarities (76–84% nucleotide identity) to miEBLN/X/P-1 (S1 Table). The flanking regions of each pair of hit sequences were well aligned between *M. natalensis* and *M. schreibersii*, indicating that they are orthologous between these species (S7 Table). We tentatively refer to these loci as miEBLN/X/P-2 and miEBLN/X/P-3. We also confirmed the presence of orthologs of these EBLN/X/Ps in the genome of *M. fuliginosus* by PCR and Sanger sequencing of genomic DNA isolated from *M. fuliginosus* (S2b Fig and S7 Table).

Next, we examined the presence of miEBLN/X/P-2 or miEBLN/X/P-3- empty loci by comparative genomics. We observed that the miEBLN/X/P-3 empty locus is present in genomes of non-miniopterid bats such as *Tadarida brasiliensis* (S2b Fig and S7 Table). We did not obtain an alignment of the flanking region of miEBLN/X/P-2 with the other bat genomes and therefore could not identify its empty locus.

We then analyzed the nucleotide sequences of miEBLN/X/P-2 and miEBLN/X/P-3. Similar to miEBLN/X/P-1, both EBLN/X/Ps contained transcription signal-like sequences, and poly-A stretches were observed downstream of T2-like sequences (S3b Fig). Furthermore, tandem duplication sequences, TGTTCA/[T/C]GTTCA and GAGGAACG/GAGGGATG, were also observed immediately upstream of S1 and downstream of T2/poly-A in miEBLN/X/P-2 and miEBLN/X/P-3, respectively (S3b Fig). These results suggest that miEBLN/X/P-2 and miEBLN/X/P-3 are also derived from ancient N/X/P transcripts and were integrated into the genome of a miniopterid bat

ancestor through LINE-1 activity.

***EBLs derived from X/P-like transcripts exist in miniopterid bat genomes***

Next, we assessed the hits that aligned to the 0.7-kb putative X/P region of miEBLN/X/P-1 (Fig 2a, teal blue lines). Two hits from *M. natalensis* and *M. schreibersii* showed 0.8 kb alignment lengths and 75% nucleotide identity to the 0.7-kb X/P region (S1 Table). The flanking regions of the hit sequences were well aligned, indicating that the loci are orthologous between these species (Fig 3b and S4a Fig; S7 Table). We also found an orthologous locus in the genome of *M. fuliginosus* using PCR and Sanger sequencing (Fig 3b and S4a Fig; S7 Table). Furthermore, we detected empty loci in genomes of non-miniopterid bats, genomes such as *Myotis lucifugus* (Fig 3b and S4a Fig; S7 Table).

The orthobornaviral X/P mRNA contains the transcription signal sequences S2/T1 in the 5'-UTR and T2 in the 3'-UTR (Fig 3c). We observed that S2/T1 and T2-like sequences are present in the 0.7-kb insertion, which are located at positions similar to those in extant X/P mRNA (Fig 3d and S5b Fig). Poly-A stretches were also observed downstream of the T2-like sequences. Furthermore, the S2 and poly-A stretches were flanked by TSDs, CTTAAGGC/CTTAAGGC (Fig 3d and S5b Fig). Taken together, these results indicated that the insertion may have been derived from an ancient X/P transcript and integrated into the genome of a miniopterid bat ancestor through LINE-1 activity. We tentatively refer to the insertion as the miEBLX/P-4.

***Preservation of intact ORFs in miEBLX/P-4 and characterization of the putatively encoded proteins***

Most EBLs do not retain viral ORFs because of the accumulation of mutations after integration. Remarkably, miEBLX/P-4 retains intact X- and P-like ORFs in different reading frames, which is consistent with the extant orthobornaviral X/P mRNA (Fig 3c). We named these ORFs as miEBLX-4 and miEBLP-4. Notably, miEBLP-4 in *M. natalensis* is disrupted by a premature stop codon.

Although there is very low sequence similarity between the amino acid sequences of these ORFs and extant orthobornaviral proteins, further analyses showed that miEBLX-4 and miEBLP-4 are intrinsically similar to extant orthobornaviral X and P proteins.

First, we examined the miEBLX-4 ORF. The ORF consists of 109 codons, which is longer than that of extant orthobornaviral X (86–90 codons). We then predicted the secondary structure and analyzed the hydrophobicity of the putative miEBLX-4 protein with respect to extant X proteins. Both extant X and miEBLX-4 were predicted to contain an alpha helix in the N-terminal region (Fig 4a and S6a Fig). The hydropathy plot showed similar hydrophobicity patterns for these proteins (S7a Fig).

We also determined whether the functional domains are conserved between extant X and miEBLX-4. Amino acid sequence alignment showed that the N-terminal region is relatively well conserved and houses functional domains previously characterized in BoDV-1 X (Fig 4c). This region includes the leucine-rich domain functioning both as a CRM1-dependent nuclear export signal (18) and a mitochondria-targeting signal (19), and is overlapped by the importin-alpha binding domain (20). The alignment also showed that leucine and arginine residues that are critical in the importin-alpha-binding

205 domain (Fig 4c; orange and blue, respectively) are conserved among extant X and
206 miEBLX-4.

207 Next, we examined the miEBLP-4 ORF. The ORF consists of 233 codons, longer
208 than those of extant orthobornaviral P (201–208 codons). The predicted secondary
209 structures and hydrophobicity pattern of miEBLP-4p were similar to those of the
210 modern counterparts (Fig 4d, S6b Fig, and S7b Fig). Alignment of the amino acid
211 sequences revealed that some of the functional domains previously characterized in
212 BoDV-1 P may be shared between extant P and miEBLP-4 (Fig 4d). Basic and proline
213 residues important for the nuclear localization signal (NLS) activity (Fig 4d, blue) (21,
214 22) in the N-terminal NLS were prominent in both proteins.

215 We also explored the interaction between miEBLX-4 and miEBLP-4, which is a
216 notable characteristic of the orthobornaviral X and P proteins (23). We expressed Myc-
217 tagged miEBLX-4 (miEBLX-4-Myc) and FLAG-tagged miEBLP-4 (miEBLP-4-FLAG)
218 in 293T cells and then immunoprecipitated them using an anti-FLAG antibody.
219 miEBLX-4-Myc was co-immunoprecipitated with miEBLP-4-FLAG, indicating the
220 interaction between these proteins (Fig 4e).

221 These results showed that miEBLX-4 and miEBLP-4 share similar intrinsic
222 properties with the extant X and P proteins. Furthermore, together with the above data,
223 they strongly suggest that miEBLN/X/P-1-3 and miEBLX/P-4 are indeed derived from
224 ancient viral N/X/P and X/P mRNAs, respectively.

225

226 ***Diverse EBLX/Ps are found in vertebrate genomes***

227 To further identify the presence of other EBLX/Ps, we conducted TFASTX and
228 TFASTY searches against vertebrate genomes using the amino acid sequences of

229 miEBLX-4 and miEBLP-4 as queries. TFASTX and TFASTY are sequence similarity
230 search programs that allow frameshifts for alignment (24). As a result, 65 unique hits
231 (S2-5 Tables) were detected in the genomes of mammals (class Mammalia), reptiles
232 (class Reptilia), amphibians (class Amphibia), and ray-finned fish (class
233 Actinopterygii). The majority of these hits were detected in the bat genomes, including
234 miEBLN/X/P-1, miEBLN/X/P-2, miEBLN/X/P-3, and miEBLX/P-4.

235 Comparative genomic analyses strongly suggest that at least some of the
236 TFASTX/Y hits are indeed derived from insertions of ancient viral X/P transcripts, as
237 follows. TFASTX/Y hits in the bat families Mormoopidae and Phyllostomidae
238 (mpEBLX/P) are orthologous, and their empty loci are present in other lineages of bats,
239 such as *Noctilio leporinus* (Fig 5a and S4b Fig). For a TFASTX/Y hit in the lone bat
240 species *Craseonycteris thonglongyai* (ctEBLX/P), we could not find its ortholog, but its
241 empty sites exist in other bat species (Fig 5b and S4c Fig). We further identified
242 orthologous TFASTX/Y hits in the rodent superfamily Octodontoidea and their empty
243 loci in other lineages of rodents (ocEBLX/P; Fig 5c and S4d Fig). We also identified
244 transcription signal-like sequences and TSDs within and flanking the above TFASTX/Y
245 hits, respectively (Fig 5d and S5b Fig). These results strongly suggested that at least
246 some of the sequences detected by TFASTX and TFASTY are bona fide EBLX/Ps.

247

248 ***Possible co-option of miEBLX/P-4 in miniopterid bats***

249 Some EVEs, including EBLs, retain intact ORFs which encode proteins that play roles
250 in host-related functions (25). As described above, miEBLX/P-4 contains intact ORFs
251 corresponding to orthobornaviral X and P genes. Therefore, we analyzed miEBLX/P-4
252 in detail.

We analyzed the expression of miEBLX/P-4 by mRNA sequencing of *M. schreibersii*-derived SuBK12-08 cells. miEBLX/P-4 is located within the intron region of *ZNF451* gene. The read coverage of the miEBLX/P-4 locus was comparable to that of *ZNF451* (Fig 6a). Interestingly, the mapping pattern revealed the existence of an alternative splicing site for *ZNF451* that could produce a chimeric transcript comprising ZNF451 exons 1–4 and the miEBLX/P-4 locus (Fig 6a).

The putative chimeric transcript contains two ORFs consisting of more than 100 codons in different reading frames: a chimeric ORF of ZNF451 in-frame with miEBLP-4 or that of the miEBLX-4 (Fig 6b). The chimeric ORF potentially encodes a fusion protein comprising the N-terminal region of ZNF451, which contains a SUMO-interacting motif (SIM) domain, and miEBLP-4 (Fig 6b).

These results suggest that miEBLX/P-4 has been co-opted by miniopterid bats as a chimeric gene with *ZNF451*.

## Discussion

We previously demonstrated that the X/P mRNA of a modern orthobornavirus can be reverse-transcribed and integrated into the host genome (3, 7); however, no EBLX and only a few EBLPs have been reported (6). We hypothesized that the technical limitations of the sequence similarity search hinder the detection of EBLs derived from the X and P genes. EBLs have been routinely searched using tBLASTn, but this can be an unsuitable method for detecting small and rapidly evolving genes such as the X and P genes (8). In this study, our aim was to develop a novel strategy to detect these hidden EBLs. To this aim, we focused on the 1.9-kb read-through transcript of orthobornaviruses, consisting of a well-conserved N gene and small and rapidly

evolving X and P genes (12), to detect hidden EBLX/Ps. We show a series of evidence implying that EBLN/X/Ps and EBLX/Ps are present in vertebrate genomes but are not detectable by BLAST searches (Figs 2 and 3). Interestingly, miEBLX/P-4 has maintained intact ORFs for more than 17 million years (Fig 3) and is expressed as a fusion transcript with the cellular gene, *ZNF451*, which potentially encodes the ZNF451/miEBLP-4 fusion protein (Fig 6). Therefore, this study revealed the presence of hidden EBLs, which could contribute to a deeper understanding of ancient bornaviruses and the co-evolution between bornaviruses and their hosts.

Here, we uncovered the potential co-option of miEBLX/P-4 by miniopterid bats, as indicated by the expression of the ZNF451/miEBLP-4 chimeric transcript, which potentially encodes a fusion protein consisting of the N-terminal region of ZNF451 and miEBLP-4 (Fig 6). To the best of our knowledge, this is the first report of a non-retroviral EVE being expressed as a chimeric transcript with a known host gene. Although we could not confirm the expression of the ZNF451/miEBLP-4 fusion protein due to the lack of antibodies, our data suggest that miEBLX/P-4 has been co-opted by miniopterid bats. Further studies are needed to understand the biological significance of miEBLX/P-4.

Our data strongly suggest that the genome organization and expression patterns of the N, X, and P genes have not substantially changed over the past million years. Here, we show that T1-read-through transcription occurred in ancient orthobornaviruses present at least 17.3 MYA (Fig 2). Furthermore, the ORF positions of the X and P genes of ancient orthobornaviruses were highly similar to those of modern orthobornaviruses (Fig 3). In addition, the sequences and order of transcription signals are highly conserved among ancient and modern orthobornaviruses (Figs 2 and 3). In BoDV-1,

301 read-through transcription has been suggested to be involved in the regulation of viral
302 gene expression (12). This hypothetical gene regulation mechanism has been
303 maintained for more than 17.3 million years. Interestingly, our previous study showed
304 that the order of S2-T1 is also conserved in some EBLNs in primates, which were
305 estimated to be integrated into the host genomes more than 43 MYA (3, 5, 9). Thus, this
306 gene regulatory mechanism may have been employed by orthobornaviruses earlier than
307 17.3 MYA.

308 The fundamental functions of viral X and P proteins may also be evolutionarily
309 conserved between ancient and modern orthobornaviruses. We observed similarities in
310 the intrinsic properties of extant orthobornaviral X and P proteins with miEBLX-4 and
311 miEBLP-4 proteins despite their very low amino acid identities (Fig 4, S6 Fig, and S7
312 Fig). Furthermore, the interaction between X and P may have been conserved in ancient
313 orthobornaviruses (Fig 4e). Therefore, although the evolutionary rates of X and P genes
314 are high, the fundamental features and functions of X and P proteins may have been
315 maintained for at least 17.3 million years.

316 For the EVE search, TFASTX/Y may be more suitable than tBLASTn, a
317 conventional sequence similarity search for the detection of EVEs. After integration into
318 the host genome, many EVEs have acquired random mutations, which caused
319 frameshifts in the ORFs. Such frameshifted EVEs are a kind of “fragmented” EVEs and
320 are thus less detectable by tBLASTn. In contrast, TFASTX/Y takes frameshifts into
321 account (24) and can thus detect such fragmented EVEs more sensitively. We are
322 currently conducting a benchmark analysis of tBLASTn and TFASTX/Y in detail,
323 showing that some of the EBLX/Ps detected by TFASTX/Y cannot be detected by
324 tBLASTn or other BLAST programs (data not shown). Therefore, TFASTX/Y can

detect fragmented EVEs more sensitively than tBLASTn, which may contribute to identifying other hidden EVEs.

The EBLX/Ps identified in this study would represent only the tip of the iceberg on the hidden EVEs that could exist in the genome. In this study, the majority of EBLX/Ps were detected in bat and rodent genomes. Note that we failed to detect any EBLX/Ps in primate genomes, which contain numerous EBLs. Considering the number of EBLs in the genomes of primates (i.e., 24 orthobornaviral EBLs were detected in the human genome (6)), it is plausible that hidden EBLX/Ps exist in primates that are undetectable because of the inherent low amino acid similarities in X and P.

We may be able to detect such hidden EBLX/Ps by identifying other EBLN/X/P integrations. However, searching for other EBLN/X/Ps by manual curation is time-consuming. Thus, the development of an efficient searching method to identify approximately 1.9-kb insertions with EBLN would facilitate further exploration of EBLX/Ps. This is probably not limited to bornaviruses; there may be many hidden EVEs that are derived from small and/or rapidly evolving viral genes. Read-through transcription has also been observed for other mononegaviruses (26). Our search strategy can be applied to the detection of such EVEs. Additionally, a previous study identified virus-like insertions using a machine learning-based method, but the origins of insertions are still unclear (27). Our approach might also contribute to understanding the origins of those sequences.

Taken together, we identified EBLs that were undetectable by conventional BLAST searches. Our findings provide novel insights into ancient bornaviruses and co-evolution between bornaviruses and their hosts. Furthermore, our search strategy will contribute to paleovirological research.

## Materials and Methods

### *Sanger sequencing of EBLs and flanking loci in M. fuliginosus*

Genomic DNA was obtained from YubFKT1 (kidney epithelial cell line) (28) in a previous study (13). miEBLN/X/P-1, miEBLN/X/P-2, miEBLN/X/P-3, and miEBLX/P-4 and corresponding flanking loci in *M. fuliginosus* were amplified by PCR using the extracted DNA with Phusion Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific). The PCR conditions are available upon request. The amplicons were purified using innuPREP PCRpure Lite Kit (Analytik Jena) and sequenced by Sanger dideoxy method at Eurofins Genomics. The primer sequences are available in S6 Table. The obtained sequences were deposited to DDBJ (accession numbers LC708263-LC708266).

### *Cell lines*

SuBK 12-08 cells (29) and human embryonic kidney 293T (HEK293T) cells were cultured in low-glucose Dulbecco's Modified Eagle Medium (Nacalai Tesque) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 100 μg/mL penicillin-streptomycin (Nacalai Tesque). The cells were maintained at 37°C in a humidified chamber containing 5% $CO_2$.

### *Construction of miEBLX-4 and miEBLP-4 expressing plasmids*

Expression vectors with C-terminal FLAG or Myc tag were first constructed. pcDNA3 was digested with EcoRI and XhoI, which was assembled with oligo DNA containing Myc or FLAG tag sequence using NEBuilder HiFi DNA Assembly Master Mix (New

373 England Biolabs). The resultant plasmids were named pcDNA3-C-Myc or pcDNA3-C-
374 FLAG. Then, miEBLX-4 and miEBLP-4 were amplified by PCR using the genomic
375 DNA extracted from YubFKT1 with Phusion Hot Start II High-Fidelity PCR Master
376 Mix, which were subsequently assembled with pcDNA3-C-Myc or pcDNA3-C-FLAG
377 linearized with KpnI and EcoRI using NEBuilder. The oligo DNA sequences are
378 available in S6 Table.
379
380 ***Co-immunoprecipitation of miEBLX-4 with miEBLP-4***
381 pcDNA3-miEBLX-4-myc and pcDNA3-miEBLP-4-FLAG plasmids (5 μg) were
382 transfected into 293T cells on 10-cm dish using 20 μL of Avalanche-Everyday
383 Transfection Reagent (EZ Biosystems) with Opti-MEM (Thermo Fisher Scientific). One
384 day post-transfection, the cells were washed twice with ice-cold PBS and lysed using
385 500 μL IP buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.4, 1 mM EDTA, 1% Triton X-
386 100, 1× protease inhibitor cocktail [Nacalai Tesque]) for 30 min at 4°C with rotation.
387 In parallel, 500 μg SureBeads Protein G magnetic beads (Bio-Rad) were incubated with
388 3 μg mouse monoclonal anti-DDDDK antibody FLA-1 (M185-3L; MBL) in IP buffer
389 for 1 hr at 4°C with rotation. The lysates were then centrifuged at 12000 × g for 10 min
390 at 4°C, and then the supernatant were immunoprecipitated with the antibody-bead
391 complex in IP buffer for 2 hr in 4°C with rotation. After incubation, the beads were
392 washed thrice with ice-cold IP buffer and then transferred into new microtubes. The
393 proteins were finally eluted from the beads by boiling in SDS sample buffer for 5 min at
394 95°C.
395

396 ***Western blotting***

397 Protein samples were separated in 12% polyacrylamide gel in Tris-glycine-SDS buffer,
398 which were transferred to PVDF membrane (MERCK) at 10 V for 1 hr using Trans-Blot
399 Turbo Transfer System (Bio-Rad). Then, the membrane was blocked with 5% skim milk
400 in 0.1% Tween 20 in PBS (PBST) for 1 hr at room temperature (RT). The membrane
401 was incubated with mouse monoclonal anti-Myc My3 (MBL; 1:5000) or anti-DDDDK
402 FLA-1 antibody (MBL; 1:2000) diluted in 5% skim milk for 1 hr at RT, washed, and
403 then incubated with horseradish peroxidase-conjugated donkey polyclonal anti-mouse
404 IgG (715-035-150, Jackson ImmunoResearch; 1:2000) antibody diluted in 5% skim
405 milk for 45 min at RT. The membrane was washed again, after which
406 chemiluminescence reaction was conducted using Chemi-Lumi One (Nacalai Tesque)
407 and was subsequently viewed using Amersham Imager 680 (GE Healthcare, IL, USA).

408

409 ***BLAST search against bat genomes to detect EBLN/X/Ps and EBLX/Ps***

410 BLASTn (14) search was conducted against the WGS database (15) of bat (Chiroptera,
411 taxid: 9397) using the nucleotide sequence of miEBLN/X/P-1 in *M. natalensis* as a
412 query on the BLAST web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Changes
413 applied to the default parameters were as follows: E-value threshold = $10^{-10}$, word size =
414 7, no filter for low complexity regions.

415

416 ***TFASTX/Y search against vertebrate genomes to detect EBLX/Ps***

417 Vertebrate genomes were downloaded using NCBI Datasets and were used as the
418 database for TFASTX/Y searches. TFASTX/Y searches, included in the FASTA36
419 package (30), were performed against each species of the vertebrate genomes using the

putative amino acid sequences miEBLX-4 and miEBLP-4 as queries with the ktup value option of 1. The obtained hits with E-values less than $10^{-4}$ were extracted, which were used for the downstream analyses.

***Identification of orthologous and syntenic empty loci***

BLASTn search was conducted against the WGS database of bats (Chiroptera, taxid: 9397) or rodents (Rodentia, taxid: 9989) using the nucleotide sequences of EBL and its flanking regions as queries on the BLAST web server. Changes applied to the default parameters were as follows: E-value = $10^{-20}$, word size = 11. The BLAST hits which comprised of genomic sequences from different species were retrieved, and then aligned by MAFFT using the E-INS-i algorithm (31) in Geneious version 11.1.5. The nucleotide accessions and region of EBLs used for search queries are available in S7 Table.

***Annotation of EBLs and TSDs***

TSD sequences were determined based on the alignment made in "*Identification of orthologous and syntenic empty loci*". To identify transcription signal-like sequences, the nucleotide sequences of EBLs and extant orthobornaviral N/X/P or X/P transcripts were aligned by MAFFT using the E-INS-I algorithm. The transcription signal-like sequences and poly-A stretches were determined according to the resultant alignment.

To identify conserved functional domains of the miEBLX-4p and miEBLP-4p, the amino acid sequences of miEBLX-4p and miEBLP-4p in *M. fuliginosus* were aligned with extant orthobornaviral X and P proteins, respectively, by MAFFT using the E-INS-i algorithm. Based on the alignments, the conserved domains were manually analyzed.

The nucleotide accession of extant orthobornaviruses used in the analysis are

available in S8 Table.

***mRNA sequencing analysis***

Total RNA was extracted from SuBK12-08 cells using NucleoSpin RNA plus (MACHEREY-NAGEL) according to the manual. Sequencing library was constructed using TruSeq Stranded mRNA Library Prep Kit (Illumina). The libraries were sequenced by NextSeq 500 using NextSeq 500/550 Mid Output Kit v2.5 (300 Cycles).

The obtained reads were preprocessed by fastp 0.23.2 (32) with the default setting, which were mapped to the reference genome of *M. natalensis* (GCF_001595765.1) by HISAT2 version 2.2.1 (33) with the default setting. The mapped reads and splicing junctions were visualized using Integrated Genomics Viewer 2.13.0 (34).

The RNA-seq data were deposited to DDBJ Sequence Read Archive (DRA) under the accession number DRR403400.

***Pairwise similarity analysis***

Amino acid sequences of the viral proteins of extant orthobornaviruses (S8 Table) were analyzed by MAFFT alignment tool using Sequence Demarcation Tool software version 1.2 (35). Minimum value for the percent identity scale was set to 0.1. The resulting matrices were set to 3-color mode and exported.

***In silico characterization of putative EBLX and EBLP proteins***

To predict the protein secondary structure, amino acid sequences of miEBLX-4 and miEBLP-4 in *M. fuliginosus*, and extant orthobornaviral X and P proteins, were analyzed by JPred version 4 (36).

To determine the protein hydropathy, the same amino acid sequences were submitted to ProtScale (37) in ExPASy Server, using the amino acid scale developed by Kyte & Doolittle. The window sizes were set to 9 and 15 for X and P proteins, respectively.

**Acknowledgments**

We are grateful to Ayato Takada for kindly providing SuBK 12-08 cell. We would like to thank Ken Maeda for kindly sharing YubFKT1. We are grateful to Takefumi Kondo and Yukari Sando for conducting the mRNA-sequencing. The animal silhouette images were downloaded under public domain access from https://commons.wikimedia.org/ and https://publicdomainvectors.org/.

This study was supported by JSPS KAKENHI grant numbers JP21H01199 (MH), 20H05682 (KT), JP21K19909 (KT), JSPS Core-to-Core Program JPJSCCA20190008 (KT), and MEXT KAKENHI grant number JP19H04833 (MH).

**References**

1. Aiewsakun P, Katzourakis A. Endogenous viruses: Connecting recent and ancient viral evolution. Virology. 2015;479-480:26-37.

2. Jern P, Coffin JM. Effects of retroviruses on host genome function. Annu Rev Genet. 2008;42:709-32.

3. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, et al. Endogenous non-retroviral RNA virus elements in mammalian genomes. Nature.

2010;463(7277):84-7.

4. Rubbenstroth D, Briese T, Durrwald R, Horie M, Hyndman TH, Kuhn JH, et al. ICTV Virus Taxonomy Profile: Bornaviridae. J Gen Virol. 2021;102(7).

5. Belyi VA, Levine AJ, Skalka AM. Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. PLoS pathogens. 2010;6(7):e1001030.

6. Kawasaki J, Kojima S, Mukai Y, Tomonaga K, Horie M. 100-My history of bornavirus infections hidden in vertebrate genomes. Proc Natl Acad Sci U S A. 2021;118(20).

7. Horie M, Kobayashi Y, Suzuki Y, Tomonaga K. Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2013;368(1626):20120499.

8. Horie M, Tomonaga K. Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses. Virus Res. 2019;262:2-9.

9. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. PLoS genetics. 2010;6(11):e1001191.

10. Gilbert C, Meik JM, Dashevsky D, Card DC, Castoe TA, Schaack S.

Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. Proc Biol Sci. 2014;281(1791):20141122.

11. Kojima S, Sato R, Yanai M, Komatsu Y, Horie M, Igarashi M, et al. Splicing-Dependent Subcellular Targeting of Borna Disease Virus Nucleoprotein Isoforms. Journal of virology. 2019;93(5).

12. Poenisch M, Wille S, Staeheli P, Schneider U. Polymerase read-through at the first transcription termination site contributes to regulation of borna disease virus gene expression. Journal of virology. 2008;82(19):9537-45.

13. Mukai Y, Horie M, Kojima S, Kawasaki J, Maeda K, Tomonaga K. An endogenous bornavirus-like nucleoprotein in miniopterid bats retains the RNA-binding properties of the original viral protein. FEBS letters. 2022;596(3):323-37.

14. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

15. Coordinators NR. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018;46(D1):D8-D13.

16. Jurka J, Klonowski P, Dagman V, Pelton P. CENSOR--a program for identification and elimination of repetitive elements from DNA sequences. Comput Chem. 1996;20(1):119-21.

17. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, et al. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. Nat Genet. 2002;31(2):159-65.

18. Ferre CA, Davezac N, Thouard A, Peyrin JM, Belenguer P, Miquel MC, et al. Manipulation of the N-terminal sequence of the Borna disease virus X protein improves its mitochondrial targeting and neuroprotective potential. FASEB J. 2016;30(4):1523-33.

19. Poenisch M, Burger N, Staeheli P, Bauer G, Schneider U. Protein X of Borna disease virus inhibits apoptosis and promotes viral persistence in the central nervous systems of newborn-infected rats. Journal of virology. 2009;83(9):4297-307.

20. Wolff T, Unterstab G, Heins G, Richt JA, Kann M. Characterization of an unusual importin alpha binding motif in the borna disease virus p10 protein that directs nuclear import. J Biol Chem. 2002;277(14):12151-7.

21. Shoya Y, Kobayashi T, Koda T, Ikuta K, Kakinuma M, Kishi M. Two proline-rich nuclear localization signals in the amino- and carboxyl-terminal regions of the Borna disease virus phosphoprotein. Journal of virology. 1998;72(12):9755-62.

22. Schwemmle M, Jehle C, Shoemaker T, Lipkin WI. Characterization of the major nuclear localization signal of the Borna disease virus phosphoprotein. J Gen

Virol. 1999;80 ( Pt 1):97-100.

23. Fujino K, Horie M, Honda T, Nakamura S, Matsumoto Y, Francischetti IM, et al. Evolutionarily conserved interaction between the phosphoproteins and X proteins of bornaviruses from different vertebrate species. PLoS One. 2012;7(12):e51161.

24. Pearson WR, Wood T, Zhang Z, Miller W. Comparison of DNA sequences with protein sequences. Genomics. 1997;46(1):24-36.

25. Horie M. The biological significance of bornavirus-derived genes in mammals. Current opinion in virology. 2017;25:1-6.

26. Banerjee AK, Barik S, De BP. Gene expression of nonsegmented negative strand RNA viruses. Pharmacology & Therapeutics. 1991;51(1):47-70.

27. Kojima S, Yoshikawa K, Ito J, Nakagawa S, Parrish NF, Horie M, et al. Virus-like insertions with sequence signatures similar to those of endogenous nonretroviral RNA viruses in the human genome. Proceedings of the National Academy of Sciences. 2021;118(5).

28. Maeda K, Hondo E, Terakawa J, Kiso Y, Nakaichi N, Endoh D, et al. Isolation of novel adenovirus from fruit bat (Pteropus dasymallus yayeyamae). Emerging infectious diseases. 2008;14(2):347-9.

29. Maruyama J, Miyamoto H, Kajihara M, Ogawa H, Maeda K, Sakoda Y, et al.

562 Characterization of the envelope glycoprotein of a novel filovirus, lloviu virus. Journal
563 of virology. 2014;88(1):99-109.
564 30. Pearson WR. Finding Protein and Nucleotide Similarities with FASTA. Curr
565 Protoc Bioinformatics. 2016;53:3 9 1-3 9 25.
566 31. Katoh K, Standley DM. MAFFT multiple sequence alignment software version
567 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80.
568 32. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ
569 preprocessor. Bioinformatics. 2018;34(17):i884-i90.
570 33. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome
571 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol.
572 2019;37(8):907-15.
573 34. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G,
574 et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
575 35. Muhire BM, Varsani A, Martin DP. SDT: a virus classification tool based on
576 pairwise sequence alignment and identity calculation. PLoS One. 2014;9(9):e108277.
577 36. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary
578 structure prediction server. Nucleic Acids Res. 2015;43(W1):W389-94.
579 37. Walker JM. The proteomics protocols handbook: Springer; 2005.

38. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol. 2017;34(7):1812-9.

## Figure legends

**Figure 1. Amino acid identities between extant orthobornaviral proteins.** Pairwise identities between the amino acid sequences of extant orthobornaviral proteins was analyzed using Sequence Demarcation Tool (35). The animal silhouettes show the representative hosts of viruses.

**Figure 2. Characterization of miEBLN-1 as an EBL derived from N/X/P transcript.** **(a)** Gene orthology analysis of the miEBLN/X/P-1 locus. Multiple alignment of the deduced syntenic loci between the indicated species is shown. Black and gray boxes show aligned regions. Thin gray lines indicate alignment gaps. Species phylogeny and minimum integration age of miEBLN/X/P-1 are indicated on the left based on TimeTree (38). The full version of this alignment is available in S1 Fig. **(b)** Schematic diagram of orthobornaviral genomic RNA, 1.9-kb N/X/P mRNA, and miEBLN/X/P-1. Transcription signals (S1, S2, T1, and T2) are shown. **(c)** Nucleotide sequence alignment of extant orthobornaviral N/X/P mRNA and miEBLN/X/P-1. Red letters indicate the start and stop codons of the N gene. Light and dark teal blue letters indicate the start codon of X gene and the stop codon of P gene, respectively. TSD, target site duplication. The full version of this alignment is available in S2 Fig.

**Figure 3. Identification of an EBL derived from X/P transcript in miniopterid bats.** **(a)** Schematic diagram of the BLASTn search against the bat whole genome shotgun sequence database, using miEBLN/X/P-1 as a query. The BLAST hits are shown by dark to light read or teal blue lines corresponding to the EBLs indicated on the right. **(b)** Gene orthology analysis of the miEBLX/P-4 locus. Species phylogeny and deduced

minimum integration age of miEBLX/P-4 are indicated on the left based on TimeTree (38). miEBLX-4 and miEBLP-4 ORFs are shown in light and dark teal blue, respectively. The full version of this alignment is available in S3 Fig. **(c)** Schematic diagram of extant orthobornaviral genomic RNA, X/P mRNA, and miEBLX/P-4. Transcription signals (S2, T1, and T2) are indicated. **(d)** Nucleotide sequence alignment of extant orthobornaviral X/P transcripts and miEBLX/P-4. Light and dark teal blue letters indicate the start codon of X gene and the stop codon of P gene, respectively. TSD, target site duplication. The full version of this alignment is available in S4 Fig.

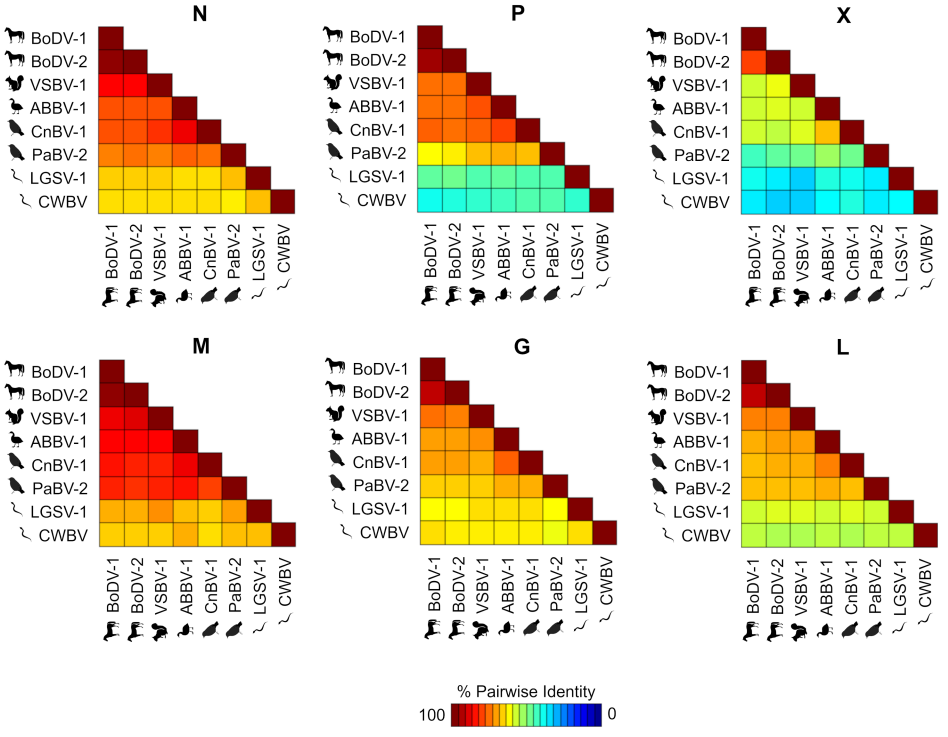
**Figure 4. Conservation of properties between putative EBLX and EBLP and viral X and P proteins. (a** and **b)** Predicted secondary structure of extant orthobornaviral X and miEBLX-4 proteins **(a)**, and extant orthobornaviral P and miEBLP-4 proteins **(b)**. H, α-helix. E, β-strand. The data including amino acid residues are shown in Figs. S5 and S6. **(c** and **d)** Amino acid sequence alignment of extant orthobornaviral X and miEBLX proteins **(c)**, and extant orthobornaviral P and miEBLP-4 proteins **(d)**. Shaded areas indicate the functional domains previously characterized in BoDV-1 X and P proteins. The conserved amino acid residues important for the functions of viral X or P protein are highlighted in colors. NLS, nuclear localization signal; NES nuclear export signal; p-Ser, phosphorylated serine residue. **(e)** Co-immunoprecipitation of Myc-tagged miEBLX-4 with FLAG-tagged miEBLP-4. FLAG-tagged miEBLP-4 was immunoprecipitated, and then analyzed by western blotting using the indicated antibodies. *, Non-specific band.

**Figure 5. Identification of diverse EBLX/Ps in the vertebrate genomes.**

**(a-c)** Gene orthology analysis of each EBLX/P locus. Species phylogeny and deduced minimum integration age of each EBLX/P are indicated on the left based on TimeTree (38). The full version of this alignment is available in S3 Fig. **(d)** Nucleotide sequence alignment of extant orthobornaviral X/P transcripts and EBLX/Ps. Black and gray boxes show aligned regions. Thin gray lines indicate alignment gaps. Light and dark teal blue letters indicate the start codon of X gene and the stop codon of P gene, respectively. TSD, target site duplication. The full version of this alignment is available in S4 Fig.

**Figure 6. Alternative splicing generates *ZNF451*/*miEBLX/P-4* chimeric transcript.** **(a)** Coverage and splicing junctions of mapped RNA reads from SuBK 12-08 cells (*M. schreibersii*) at the *ZNF451* locus where miEBLX/P-4 is located. The miEBLX/P-4 region, mapped read coverage graph on miEBLX/P-4, and splicing junction between ZNF451 and miEBLX/P-4 are shown in teal blue. **(b)** Schematic figure of putative chimeric ZNF451/miEBLP-4 protein. Shaded teal blue box SIM, SUMO interacting motif.

**Fig. 1**


N
P
X
M
G
L
BoDV-1
BoDV-2
VSBV-1
ABBV-1
CnBV-1
PaBV-2
LGSV-1
CWBV
% Pairwise Identity
100
0

# Fig. 2

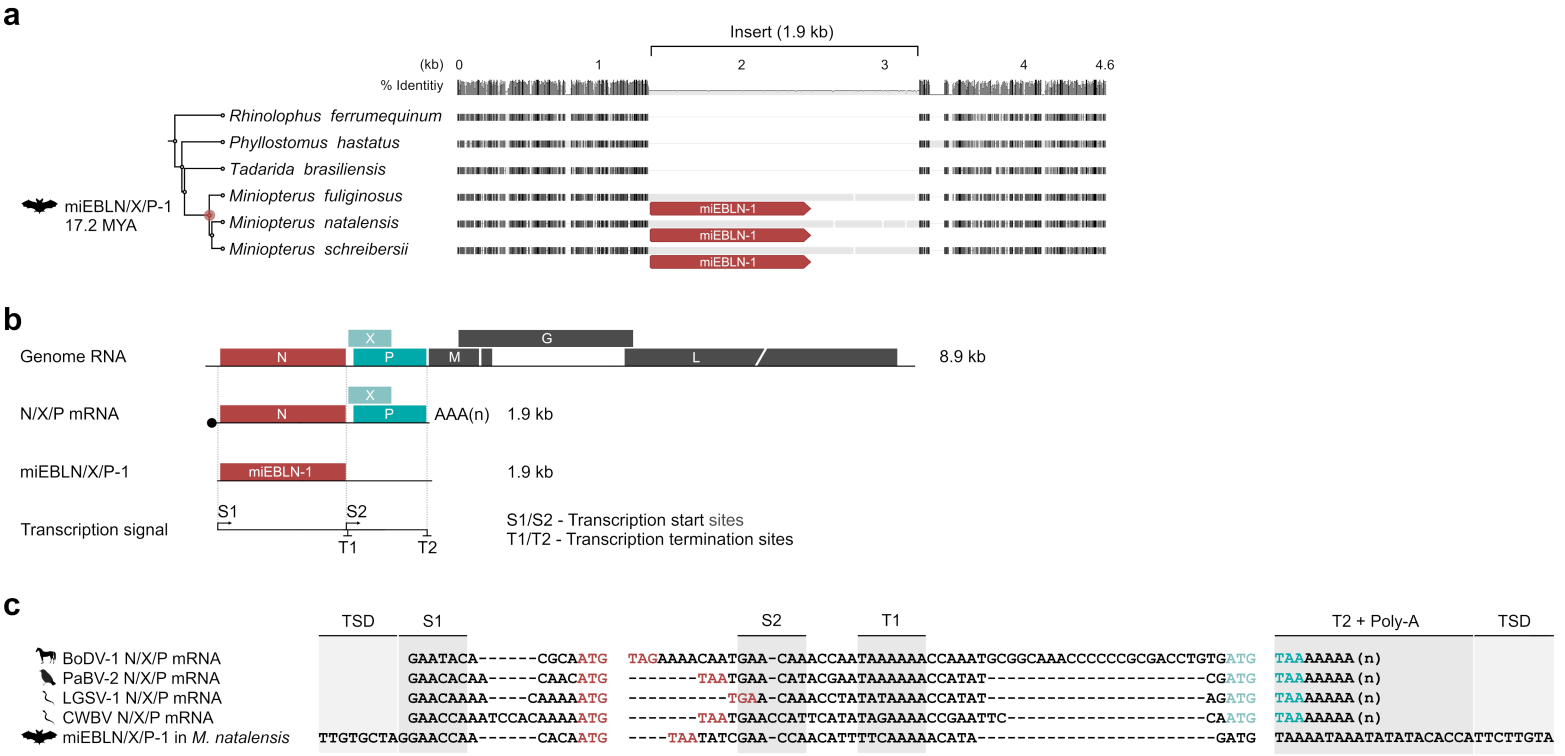

a
Insert (1.9 kb)
(kb) 0 1 2 3 4 4.6
% Identitiy
Rhinolophus ferrumequinum
Phyllostomus hastatus
Tadarida brasiliensis
Miniopterus fuliginosus
Miniopterus natalensis
Miniopterus schreibersii
miEBLN/X/P-1
17.2 MYA
miEBLN-1
b
Genome RNA
N
X
P
M
G
L
8.9 kb
N/X/P mRNA
AAA(n)
1.9 kb
miEBLN/X/P-1
1.9 kb
Transcription signal
S1
S2
T1
T2
S1/S2 - Transcription start sites
T1/T2 - Transcription termination sites
c
TSD
S1
S2
T1
T2 + Poly-A
TSD
BoDV-1 N/X/P mRNA
PaBV-2 N/X/P mRNA
LGSV-1 N/X/P mRNA
CWBV N/X/P mRNA
miEBLN/X/P-1 in M. natalensis

Fig. 3

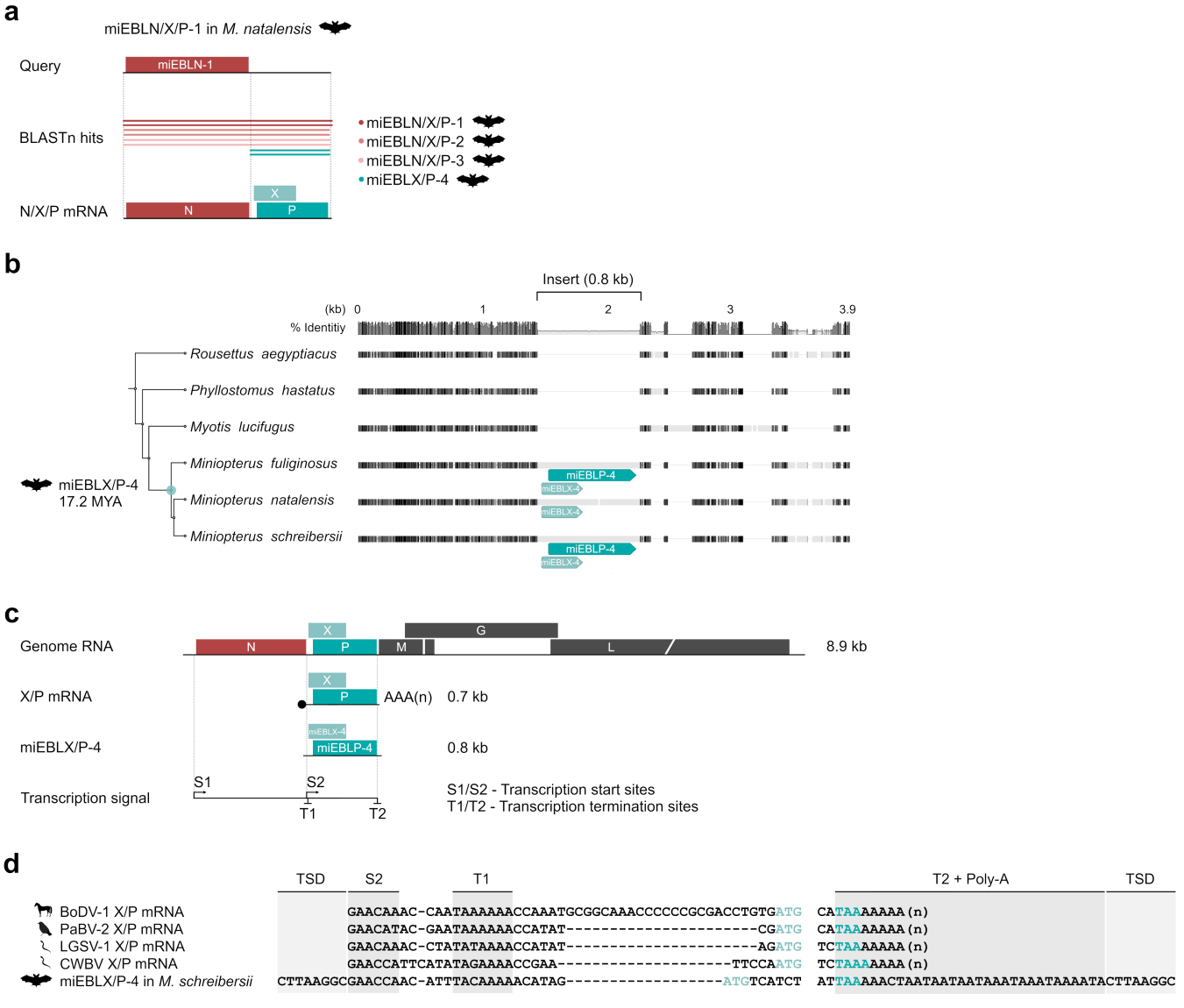

a
miEBLN/X/P-1 in M. natalensis
Query
miEBLN-1
BLASTn hits
N/X/P mRNA
N
X
P
miEBLN/X/P-1
miEBLN/X/P-2
miEBLN/X/P-3
miEBLX/P-4
b
Insert (0.8 kb)
(kb) 0 1 2 3 3.9
% Identitiy
Rousettus aegyptiacus
Phyllostomus hastatus
Myotis lucifugus
Miniopterus fuliginosus
Miniopterus natalensis
Miniopterus schreibersii
miEBLX/P-4
17.2 MYA
miEBLP-4
miEBLX-4
c
Genome RNA
N
X
P
M
G
L
8.9 kb
X/P mRNA
AAA(n)
0.7 kb
miEBLX/P-4
miEBLP-4
0.8 kb
Transcription signal
S1
S2
T1
T2
S1/S2 - Transcription start sites
T1/T2 - Transcription termination sites
d
TSD
S2
T1
T2 + Poly-A
TSD
BoDV-1 X/P mRNA
PaBV-2 X/P mRNA
LGSV-1 X/P mRNA
CWBV X/P mRNA
miEBLX/P-4 in M. schreibersii
GAACAAAC-CAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTGATG CATAAAAAAA(n)
GAACATAC-GAATAAAAAACCATAT----------CGATG CATAAAAAAA(n)
GAACAAAC-CTATATAAAACCATAT----------AGATG TCTAAAAAAA(n)
GAACCATTCATATAGAAAACCGAA----------TTCCAATG TCTAAAAAAA(n)
CTTAAGGCGAACCAAC-ATTTACAAAAACATAG----------ATGTCATCT ATTAAAAACTAATAATAATAAATAAATAAAATACTTAAGGC

**Fig. 4**

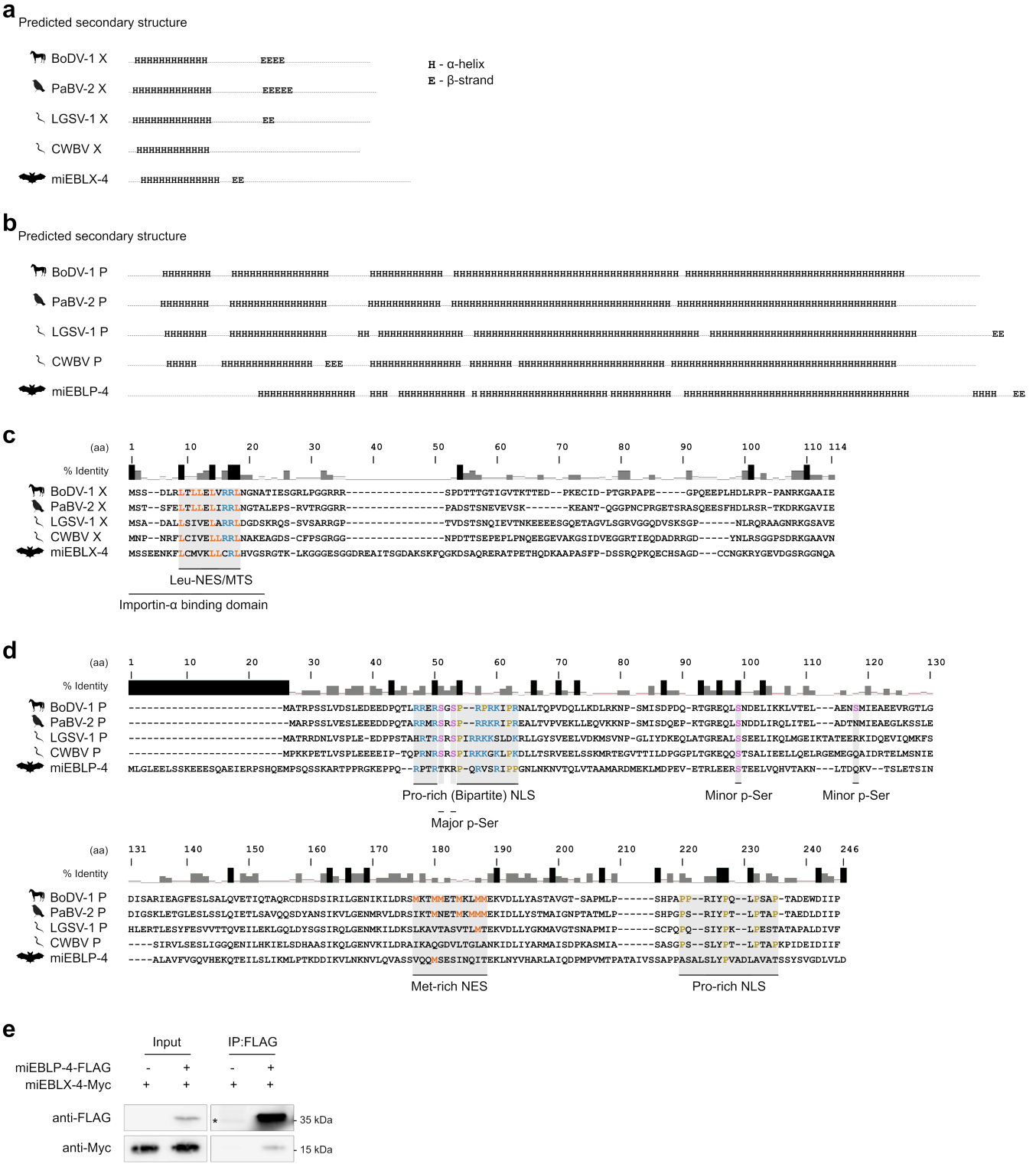
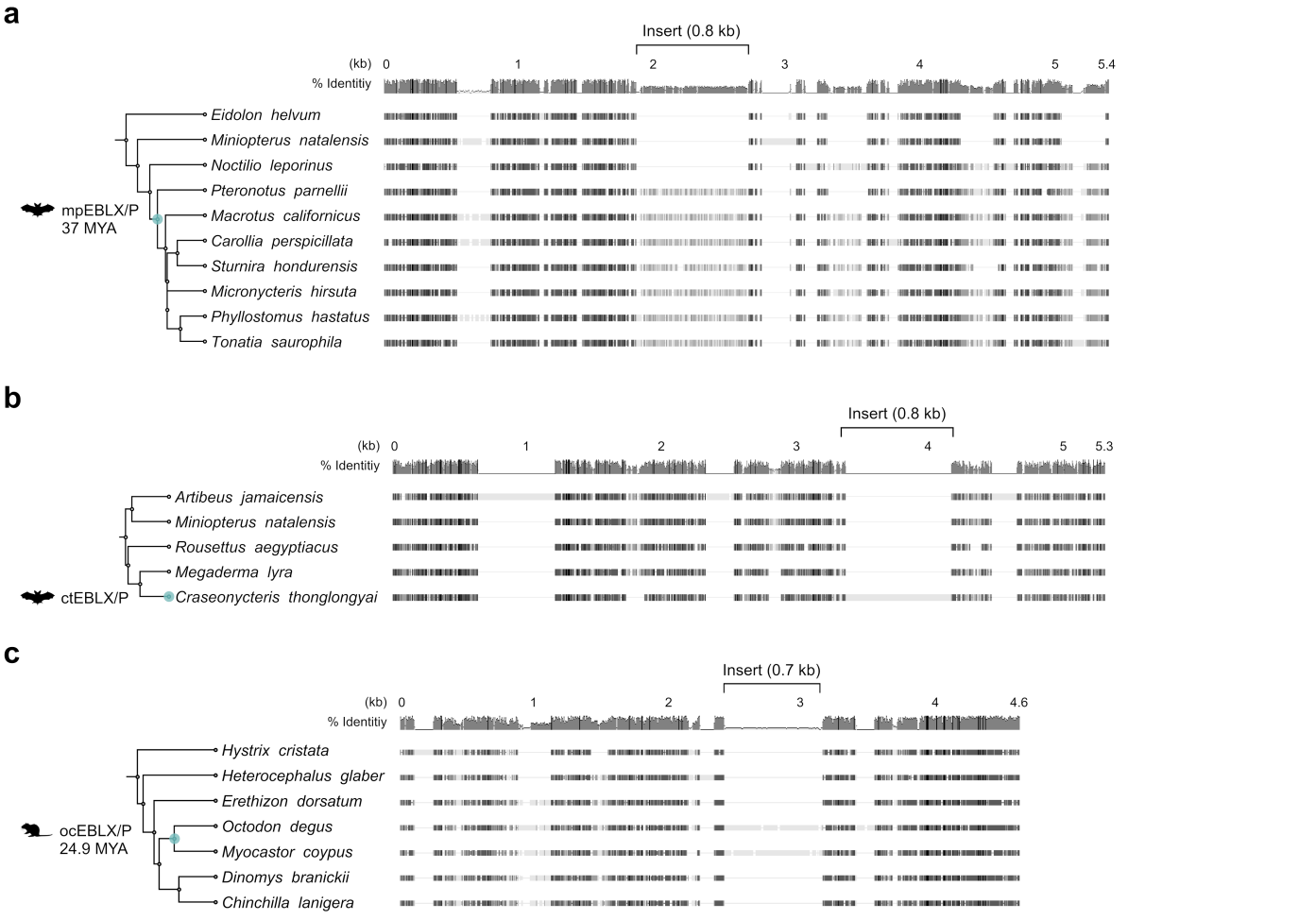

a
Predicted secondary structure
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX-4
H - α-helix
E - β-strand
b
Predicted secondary structure
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP-4
c
(aa)
% Identity
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX-4
Leu-NES/MTS
Importin-α binding domain
d
(aa)
% Identity
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP-4
Pro-rich (Bipartite) NLS
Minor p-Ser
Minor p-Ser
Major p-Ser
Met-rich NES
Pro-rich NLS
e
Input
IP:FLAG
miEBLP-4-FLAG
miEBLX-4-Myc
anti-FLAG
anti-Myc
35 kDa
15 kDa

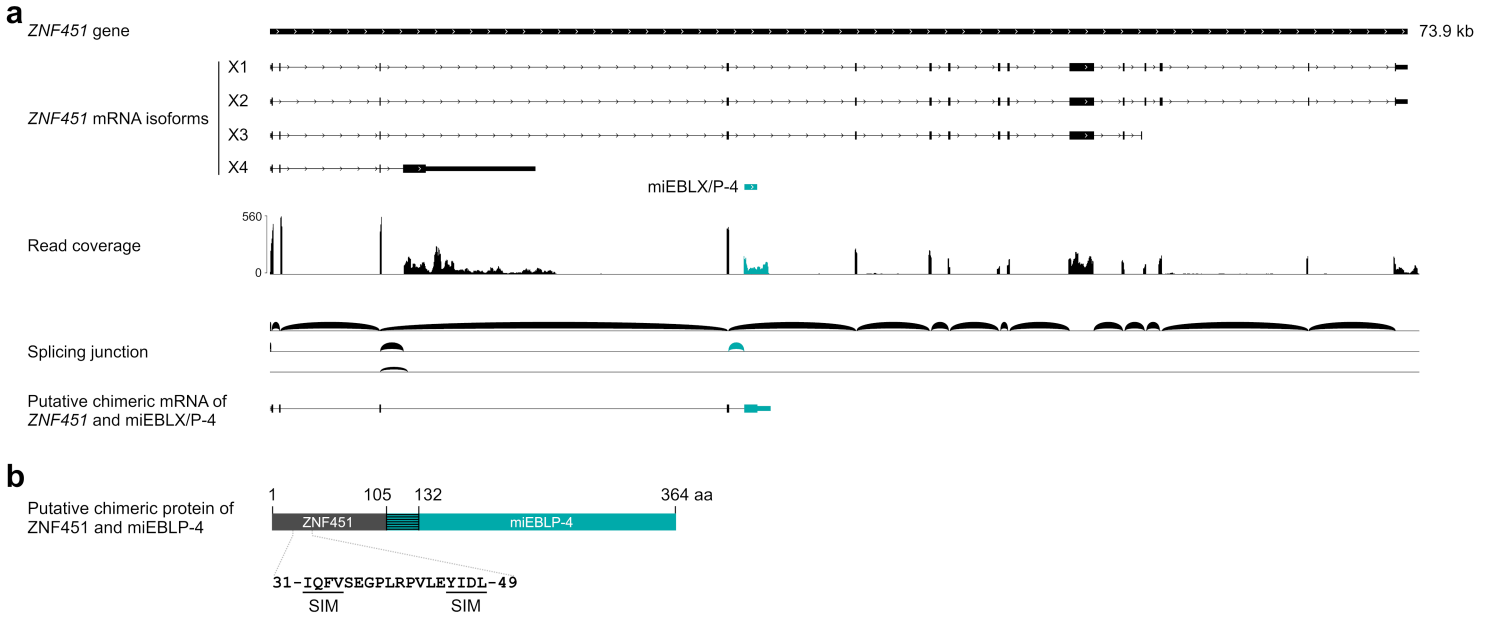
**Fig. 5**




**d**

| | TSD | S2 | T1 | T2 + Poly-A | TSD |
|---|---|---|---|---|---|
| BoDV-1 X/P mRNA | | GAACAAAC- | CAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTGATG | CATAAAAAAA(n) | |
| PaBV-2 X/P mRNA | | GAACATAC- | GAATAAAAAACCATAT----------CGATG | CATAAAAAAA(n) | |
| LGSV-1 X/P mRNA | | GAACAAAC- | CTATATAAAACCATAT----------AGATG | TCTAAAAAAA(n) | |
| CWBV X/P mRNA | | GAACCATTCATATAGAAAACCGAA- | ----------TTCCAATG | TCTAAAAAAA(n) | |
| mpEBLX/P in *P. parnellii* | GAACTTGGCTTTTGGAACCTGT- | CTATATAAAAACATAG- | ----------GTTTACATG | TTAAAAAAAAAAAAAAAAA- | ------GAACTTGGCTTTCA |
| ctEBLX/P in *C. thonglongyai* | TACAGGCATTGAGGAGCCAAC- | ATATGTAAAAACATAG- | ----------ATGTCTGTT | GTTAAAAAAAA- | ----------TACAGACATTGAC |
| ocEBLX/P in *O. degus* | AATAGGAACCAAGCCACTAGAAAACCATAC- | | ----------CCGTC | GCTAAAAAAATAAAATAAAATAAATAAATA | |

Fig. 6


a
ZNF451 gene
73.9 kb
ZNF451 mRNA isoforms
X1
X2
X3
X4
miEBLX/P-4
Read coverage
560
0
Splicing junction
Putative chimeric mRNA of ZNF451 and miEBLX/P-4
b
Putative chimeric protein of ZNF451 and miEBLP-4
1
105
132
364 aa
ZNF451
miEBLP-4
31-IQFVSEGPLRPVLEYIDL-49
SIM
SIM