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], & Masayuki Horie [1,4*]

[1] Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, Izumisano 598-8531, Osaka, Japan

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

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

[4] Osaka International Research Center for Infectious Diseases, Osaka Metropolitan University, Izumisano 598-8531, Osaka, Japan

***Corresponding author**

Masayuki Horie, DVM, PhD

Tel: +81-72-463-5694

Email: mhorie@omu.ac.jp

Keywords: endogenous viral element, bornavirus, orthobornavirus, paleovirology, readthrough transcript

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, our aim was to develop a novel strategy to detect such "hidden" EBLs. To that aim, we focused on the 1.9 kb readthrough 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 ZNF451EBLP-4 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 [2]. As molecular fossils of ancient viruses, EVEs have expanded our knowledge of the diversity, host range, and long-term evolution of viruses. 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, 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, 6-8]. 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, our 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 [9], 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 readthrough 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 [10]. 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 Fig. S1a). 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.

We carried out a series of analyses to understand the source of the 0.7-kb unannotated region. First, we performed a BLASTn search against the NCBI nt database 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 [11], 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. 2b), which led us to speculate that the insertion may be derived from an N/X/P transcript. N/X/P mRNA is produced by

readthrough 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). 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 Fig. S2). Furthermore, a tandem duplication sequence, TTGTGCTA/TTCTTGTA, was observed immediately upstream and downstream of S1 and T2-poly-A sequences, respectively. This is a potential target site duplication (TSD), which is a hallmark of LINE-1-mediated mRNA integration [13]. 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). 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 (Fig. 3a). We obtained six hits from the genomes of *M. natalensis* and *M. schreibersii*. While four hits were aligned in the entire N/X/P region of miEBLN/X/P-1, interestingly, two other hits were aligned only to the putative X/P region (Fig. 3a and Table S1).

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 (Table S1). 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 (Table S7). 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* (Table S7).

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* (Fig. S1B and Table S7). 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 (Fig. S2). 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. 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 aligned to the 0.7-kb putative X/P region of miEBLN/X/P-1.

166 Two hits from *M. natalensis* and *M. schreibersii* showed 0.8 kb alignment lengths and
167 75% nucleotide identity to the 0.7-kb X/P region (Table S1). The flanking regions of the
168 hit sequences were well aligned, indicating that the loci are orthologous between these
169 species (Table S7). We also found an orthologous locus in the genome of *M. fuliginosus*
170 using PCR and Sanger sequencing (Table S7). Furthermore, we detected empty loci in
171 genomes of non-miniopterid bats, genomes such as *Myotis lucifugus* (Fig. 3b and Fig.
172 S3a).

173 The orthobornaviral X/P mRNA contains the transcription signal sequences S2/T1
174 in the 5'-UTR and T2 in the 3'-UTR. We observed that S2/T1 and T2-like sequences are
175 present and located at positions similar to those in extant X/P mRNA. Additionally,
176 poly-A stretches were observed downstream of the T2 sequences. Furthermore, the S2
177 and poly-A stretches were flanked by TSDs, CTTAAGGC/CTTAAGGC (Fig. 3a, b and
178 Fig. S5). Taken together, these results indicated that the locus may have been derived
179 from an ancient X/P transcript and integrated into the genome of a miniopterid bat
180 ancestor through LINE-1 activity. We tentatively refer to it as the miEBLX/P-4.

181

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

184 Most EBLs do not retain viral ORFs because of the accumulation of mutations after
185 integration. Remarkably, however, miEBLX/P-4 retains intact X- and P-like ORFs in
186 different reading frames, which is consistent with the extant orthobornaviral X/P mRNA
187 (Fig. 3c and Fig. S4a). Therefore, we named these ORFs as miEBLX-4 and miEBLP-4.
188 Notably, miEBLP-4 in *M. natalensis* is disrupted by a premature stop codon.

189 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 the 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 Fig. S6a). The hydropathy plot showed similar hydrophobicity patterns for these proteins (Fig. S7a).

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 [14] and a mitochondria-targeting signal [15], and is overlapped by the importin-alpha binding domain [16]. The alignment also showed that leucine and arginine residues that are critical in the importin-alpha-binding domain (Fig. 4d; orange and blue, respectively) are conserved among extant X and miEBLX-4.

Next, we examined the miEBLP-4 ORF. The ORF consists of 233 codons; it is therefore longer than those of extant orthobornaviral P (201–208 codons). Extant P and putative miEBLP-4 proteins also showed similar predicted secondary structures (Fig. 4b and Fig. S6b) and hydrophobicity patterns (Fig. S7b). The predicted secondary structures of miEBLP-4p were similar to those of the modern counterparts (Fig. 4b and Fig. S6b). Alignment of the amino acid sequences revealed that some of the functional

214 domains previously characterized in BoDV-1 P may be shared between extant P and
215 miEBLP-4 (Fig. 4d). Proline residues (Fig. 4d, blue) in the N-terminal and C-terminal
216 nuclear localization signals were prominent in both proteins [17].

217 We also explored the interaction between miEBLX-4 and miEBLP-4, which is a
218 notable characteristic of the orthobornaviral X and P proteins [18]. We expressed Myc-
219 tagged miEBLX-4 (miEBLX-4-Myc) and FLAG-tagged miEBLP-4 (miEBLP-4-FLAG)
220 in 293T cells and then immunoprecipitated them using an anti-FLAG antibody.
221 miEBLX-4-Myc was co-immunoprecipitated with miEBLP-4-FLAG, indicating the
222 interaction between these proteins (Fig. 4e).

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

227

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

229 To further identify the presence of other EBLX/Ps, we conducted TFASTX and
230 TFASTY searches against vertebrate genomes using the amino acid sequences of
231 miEBLX-4 and miEBLP-4 as queries. TFASTX and TFASTY are sequence similarity
232 search programs that allow frameshifts for alignment [19]. As a result, 65 unique hits
233 (Table S2–S5) were detected in the genomes of mammals (class Mammalia), reptiles
234 (class Reptilia), amphibians (class Amphibia), and ray-finned fish (class
235 Actinopterygii). The majority of these hits were detected in the bat genomes, including
236 miEBLN/X/P-1, miEBLN/X/P-2, miEBLN/X/P-3, and miEBLX/P-4.

237 Comparative genomic analyses strongly suggest that at least some of the

238 TFASTX/Y hits are indeed derived from insertions of ancient viral X/P transcripts, as
239 follows. TFASTX/Y hits in the bat families Mormoopidae and Phyllostomidae
240 (mpEBLX/P) are orthologous, and their empty loci are present in other lineages of bats,
241 such as *Noctilio leporinus* (Fig. 5a and Fig. S3b). For a TFASTX/Y hit in the lone bat
242 species *Craseonycteris thonglongyai* (ctEBLX/P), we could not find its ortholog, but its
243 empty sites exist in other bats (Fig. 5b and Fig. S3c). We further identified orthologous
244 TFASTX/Y hits in the rodent superfamily Octodontoidea and their empty loci in other
245 lineages of rodents (ocEBLX/P; Fig. 5c and Fig. S3d). We also identified transcription
246 signal-like sequences and TSDs within and flanking the above TFASTX/Y hits,
247 respectively (Fig. 5d and Fig. S4). These results strongly suggested that at least some of
248 the sequences detected by TFASTX and TFASTY are bona fide EBLX/Ps.

249 Interestingly, the insertion sizes of bat mpEBLX/P and ctEBLX/P were both 0.8
250 kb, which is similar to that of miEBLX/P-4. In contrast, that of rodent ctEBLX/P is 0.7
251 kb, which is closer to the size of extant orthobornaviral X/P transcripts (see discussion).

252

253 ***Possible co-option of EBLX/P in miniopterid bats***

254 Some EVEs, including EBLs, retain intact ORFs that encode proteins that play roles in
255 host-related functions [20]. As described above, miEBLX/P-4 contains intact ORFs
256 corresponding to orthobornaviral X and P genes. Therefore, we analyzed miEBLX/P-4
257 in detail. We first examined the location of miEBLX/P-4 in the *M. natalensis* genome.
258 miEBLX/P-4 is located within the intron region of *ZNF451* gene (Fig. 6).

259 Next, we analyzed the expression of miEBLX/P-4 by mRNA sequencing of *M.*
260 *schreibersii*-derived SubK12-08 cells. The read coverage of the miEBLX/P-4 locus was
261 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. The chimeric ORF potentially encodes a fusion protein comprising the N-terminal region of ZNF451, which contains a SUMO-interacting motif (SIM) domain, a short intronic region of *ZNF451*, and miEBLP-4 (Fig. 6).

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 [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 that aim, we focused on the 1.9-kb readthrough 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. Interestingly, miEBLX/P-4 has maintained intact ORFs for more than 17 million years and is expressed as a fusion transcript with the cellular

gene, *ZNF451*, which potentially encodes the ZNF451/miEBLP-4 fusion protein. 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. 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 co-opted by miniopterid bats. Further studies are needed to understand the biological significance of EBLs.

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-readthrough 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, readthrough transcription has been suggested to be involved in the regulation of viral gene expression [12]. This hypothetical gene regulation mechanism has been maintained for more than 17.3 million years. Interestingly, our previous study showed that the order of S2-T1 is also conserved in some EBLNs shared among Old World monkeys, which were estimated to be integrated into the host genomes more than 43

310 MYA [3]. Thus, this gene regulatory mechanism may have been employed by
311 orthobornaviruses earlier than 17.3 MYA.

312 The fundamental functions of viral X and P proteins may also be evolutionarily
313 conserved between ancient and modern orthobornaviruses. We observed similarities in
314 the intrinsic properties of extant orthobornaviral X and P proteins with miEBLX-4p and
315 miEBLP-4p despite their very low amino acid identities (Fig. 4). Furthermore, the
316 interaction between X and P may have been conserved in ancient orthobornaviruses
317 (Fig. 4). Although the evolutionary rates of X and P genes are high, the fundamental
318 features and functions of X and P proteins may have been maintained for at least 17.3
319 million years. Some orthobornaviruses are known to be fetal zoonotic pathogens [21].
320 Thus, although the biological functions of the X protein and the interaction between X
321 and P proteins are not fully understood, these proteins can be good targets for the
322 development of antivirals.

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

333 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. Readthrough transcription has also been observed for other mononegaviruses. Our search strategy can be applied to the detection of such EVEs.

Taken together, we identified EBLs that were undetectable by conventional BLAST searches. Our findings provide novel insights into ancient bornaviruses. 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) [22] in a previous study [10]. miEBLN/X/P-1, miEBLN/X/P-2, miEBLN/X/P-3, and miEBLX/P-4 and corresponding flanking locus in *M. fuliginosus* were PCR-amplified from the DNA extracted from YubFKT1 using Phusion Hot Start II High-Fidelity PCR Master

358 Mix (Thermo Fisher Scientific). The amplicons were then purified using innuPREP
359 PCRpure Lite Kit (Analytik Jena) and sequenced by Sanger dideoxy method at Eurofins
360 Genomics. The primer sequences are available in Table S6. The obtained sequences
361 were deposited to DDBJ (accession numbers LC708263-LC708266).
362
363 ***Cell lines***
364 SuBK 12-08 cells [23] and human embryonic kidney 293T (HEK293T) cells were
365 cultured in low-glucose Dulbecco's Modified Eagle Medium (Nacalai Tesque)
366 supplemented with 10% fetal bovine serum (Sigma-Aldrich) and 100 μg/mL penicillin-
367 streptomycin (Nacalai Tesque). The cells were then maintained at 37°C in a humidified
368 chamber containing 5% CO_2.
369
370 ***Construction of miEBLX-4 and miEBLP-4 expressing plasmids***
371 Expression vectors with C-terminal FLAG or Myc tag were first constructed. pcDNA3
372 was digested with EcoRI and XhoI, which was assembled with oligo DNA containing
373 Myc or FLAG tag sequence using NEBuilder HiFi DNA Assembly Cloning Kit (New
374 England Biolabs). The resultant plasmids were named pcDNA3-C-Myc or pcDNA3-C-
375 FLAG. miEBLX-4 and miEBLP-4 sequences were then PCR-amplified from the
376 YubFKT1 DNA sample, and were subsequently assembled with pcDNA3-C-Myc or
377 pcDNA3-C-FLAG linearized with KpnI and EcoRI using NEBuilder. The oligo DNA
378 sequences are available in Table S6.
379
380 ***Co-immunoprecipitation of miEBLX-4 with miEBLP-4***
381 The constructed plasmids (5 μg) were transfected into seeded 293T cells on 10-cm dish

using 20 μL of Avalanche-Everyday Transfection Reagent (EZ Biosystems), with Opti-MEM (Thermo Fisher Scientific) as the transfecting medium. One day post-transfection, the cells were washed twice with ice-cold PBS and lysed using 500 μL IP buffer (150 mM NaCl, 20 mM Tris-HCl pH 7.4, 1 mM EDTA, 1% Triton X-100, 1x protease inhibitor cocktail) for 30 min in 4°C with rotation. In parallel, 500 μg SureBeads Protein G magnetic beads (Bio-Rad) were incubated with 3 μg mouse monoclonal anti-DDDDK antibody FLA-1 (M185-3L; MBL) in IP buffer for 1 hr in 4°C with rotation. The lysates were then centrifuged at 12000 x g for 10 min at 4°C, after which the FLAT-tagged proteins were immunoprecipitated from the lysate supernatant by mixing with the antibody-bead complex in IP buffer for 2 hr in 4°C with rotation. After incubation, the beads were washed thrice with ice-cold IP buffer and then transferred into new microtubes. The proteins were finally eluted from the beads by boiling in SDS sample buffer (組成) for 5 min in 95°C.

Western blotting

Protein samples were loaded in 12% polyacrylamide gel and run under electrophoresis in Tris-glycine-SDS buffer. The proteins were then semi-dry transferred to PVDF membrane at 10 V for 1 hr. Following blot transfer, the membrane was blocked with 5% skim milk in 0.1% Tween 20 in PBS (PBST) for 1 hr at room temperature (RT). The membrane was incubated with mouse monoclonal anti-Myc My3 (M192-3; MBL) or anti-DDDDK antibody diluted in 5% skim milk for 1 hr at RT, washed, and then incubated with horseradish peroxidase-conjugated donkey polyclonal anti-mouse IgG (715-035-150; Jackson ImmunoResearch) antibody diluted in 5% skim milk for 45 min

405 at RT. The membrane was washed again, after which chemiluminescence reaction was
406 conducted using Chemi-Lumi One (Nacalai Tesque) and was subsequently viewed using
407 Amersham Imager 680 (GE Healthcare, IL, USA).
408
409 ***BLAST search against bat genomes to detect EBLN/X/Ps and EBLX/Ps***
410 BLASTn [24] search was conducted using the nucleotide sequence of miEBLN/X/P-1
411 in *M. natalensis* as query, against the WGS database [25] of bat (Chiroptera, taxid:
412 9397) on the BLAST web server (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Changes
413 applied to default parameters were as follows: E-value threshold = 10^{-10}, word size = 7,
414 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 search. TFASTX/Y searches, included in FASTA36 package
419 [26], were performed against each vertebrate genome using the putative amino acid
420 sequences miEBLX-4 and miEBLP-4 with the ktup value of 1.
421
422 ***Identification of orthologous and empty loci***
423 BLASTn search was conducted using the nucleotide sequences of EBL and its flanking
424 regions as query, against the WGS database of bats (Chiroptera, taxid: 9397) or rodents
425 (Rodentia, taxid: 9989) on the BLAST web server. Changes applied to default
426 parameters were as follows: E-value = 10^{-20}, word size = 11. The BLAST hits which
427 comprised of genomic sequences from different species were retrieved, and then aligned
428 by MAFFT using the E-INS-i algorithm [27] in Geneious version 11.1.5. The nucleotide

accessions and loci region of EBLs used for search query are available in Table S7.

Annotation of EBLs and their TSDs

TSD sequences were determined based on the above loci alignment. 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 then determined according to the resulting 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. The conserved domains were manually analyzed based on the previously characterized functional domains in BoDV-1 X and P and the alignments.

The nucleotide accession of extant orthobornaviruses used in the analysis are available in Table S8. Multiple sequence alignments and annotations were conducted using Geneious version 11.1.5.

mRNA-seq analysis

Total RNA sample from Subk12-08 cells was obtained by 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 [28] with the default setting, which were mapped to the reference genome of *M. natalensis* (GCF_001595765.1) by

HISAT2 version 2.2.1 [29] with the default setting ref. The mapped reads and splicing junctions were visualized using Integrated Genomics Viewer 2.13.0 [30].

Pairwise similarity analysis

Amino acid sequences of the viral proteins of extant orthobornaviruses (Table S8) were analyzed by MAFFT alignment tool using Sequence Demarcation Tool software version 1.2 [31]. 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 submitted to JPred version 4 [32]. The simple results were then exported.

To determine the protein hydrophathy, the same amino acid sequences were submitted to ProtScale [33] 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. The hydropathy scores were then exported.

Acknowledgements

This study was supported by JSPS KAKENHI grant numbers JP21H01199 (MH) and 20H05682 (KT) and MEXT KAKENHI grant number JP19H04833 (MH).

The animal silhouette images were downloaded under public domain access from https://commons.wikimedia.org/ and https://publicdomainvectors.org/.

Conflict of interest

None declared.

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**:709-732.
3. Horie M, Honda T, Suzuki Y, Kobayashi Y, Daito T, Oshida T, Ikuta K, Jern P, Gojobori T, Coffin JM *et al*: **Endogenous non-retroviral RNA virus elements in mammalian genomes**. *Nature* 2010, **463**(7277):84-87.
4. Rubbenstroth D, Briese T, Durrwald R, Horie M, Hyndman TH, Kuhn JH, Nowotny N, Payne S, Stenglein MD, Tomonaga K *et al*: **ICTV Virus Taxonomy Profile: Bornaviridae**. *The Journal of general virology* 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 Pathog* 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 research* 2019, **262**:2-9.
9. Kojima S, Sato R, Yanai M, Komatsu Y, Horie M, Igarashi M, Tomonaga K: **Splicing-Dependent Subcellular Targeting of Borna Disease Virus Nucleoprotein Isoforms**. *Journal of virology* 2019, **93**(5).
10. 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):323-337.

11. 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):119-121.
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):9537-9545.
13. Morrish TA, Gilbert N, Myers JS, Vincent BJ, Stamato TD, Taccioli GE, Batzer MA, Moran JV: **DNA repair mediated by endonuclease-independent LINE-1 retrotransposition**. *Nat Genet* 2002, **31**(2):159-165.
14. Ferre CA, Davezac N, Thouard A, Peyrin JM, Belenguer P, Miquel MC, Gonzalez-Dunia D, Szelechowski M: **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):1523-1533.
15. 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):4297-4307.
16. 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):12151-12157.
17. 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):9755-9762.
18. Fujino K, Horie M, Honda T, Nakamura S, Matsumoto Y, Francischetti IM, Tomonaga K: **Evolutionarily conserved interaction between the phosphoproteins and X proteins of bornaviruses from different vertebrate species**. *PloS one* 2012, **7**(12):e51161.
19. Pearson WR, Wood T, Zhang Z, Miller W: **Comparison of DNA sequences with protein sequences**. *Genomics* 1997, **46**(1):24-36.
20. Horie M: **The biological significance of bornavirus-derived genes in mammals**. *Current opinion in virology* 2017, **25**:1-6.
21. Rubbenstroth D, Schlottau K, Schwemmle M, Rissland J, Beer M: **Human bornavirus research: Back on track!** *PLoS Pathog* 2019, **15**(8):e1007873.
22. Maeda K, Hondo E, Terakawa J, Kiso Y, Nakaichi N, Endoh D, Sakai K, Morikawa S, Mizutani T: **Isolation of novel adenovirus from fruit bat**

(Pteropus dasymallus yayeyamae). *Emerging infectious diseases* 2008, **14**(2):347-349.

23. Maruyama J, Miyamoto H, Kajihara M, Ogawa H, Maeda K, Sakoda Y, Yoshida R, Takada A: **Characterization of the envelope glycoprotein of a novel filovirus, lloviu virus**. *Journal of virology* 2014, **88**(1):99-109.
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC Bioinformatics* 2009, **10**:421.
25. Coordinators NR: **Database resources of the National Center for Biotechnology Information**. *Nucleic Acids Res* 2018, **46**(D1):D8-D13.
26. Pearson WR: **Finding Protein and Nucleotide Similarities with FASTA**. *Curr Protoc Bioinformatics* 2016, **53**:3 9 1-3 9 25.
27. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Mol Biol Evol* 2013, **30**(4):772-780.
28. Chen S, Zhou Y, Chen Y, Gu J: **fastp: an ultra-fast all-in-one FASTQ preprocessor**. *Bioinformatics* 2018, **34**(17):i884-i890.
29. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL: **Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype**. *Nat Biotechnol* 2019, **37**(8):907-915.
30. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: **Integrative genomics viewer**. *Nat Biotechnol* 2011, **29**(1):24-26.
31. .
32. Drozdetskiy A, Cole C, Procter J, Barton GJ: **JPred4: a protein secondary structure prediction server**. *Nucleic Acids Res* 2015, **43**(W1):W389-394.
33. Walker JM: **The proteomics protocols handbook**: Springer; 2005.

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 ref.

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 gaps. Species phylogeny and minimum integration age of miEBLN/X/P-1 are indicated on the left based on TimeTree ref. **(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.

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 query. The BLAST hits are shown by dark to light read or teal 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. **(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 letters indicate the start codon of X gene and the stop codon of P gene, respectively. TSD, target site duplication.

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. **(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. **(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. **(D)** Nucleotide sequence alignment of extant orthobornaviral X/P transcripts and EBLX/Ps. Light teal and dark teal letters indicate the start codon of X gene and the stop codon of P gene, respectively. TSD, target site duplication.

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. **(B)** Schematic diagram of putative ZNF451/EBLX/P-4 chimeric transcript. The transcript contains two ORFs in different frames: miEBLX-4 alone and ZNF451 fused with miEBLP-4.

Fig. 1

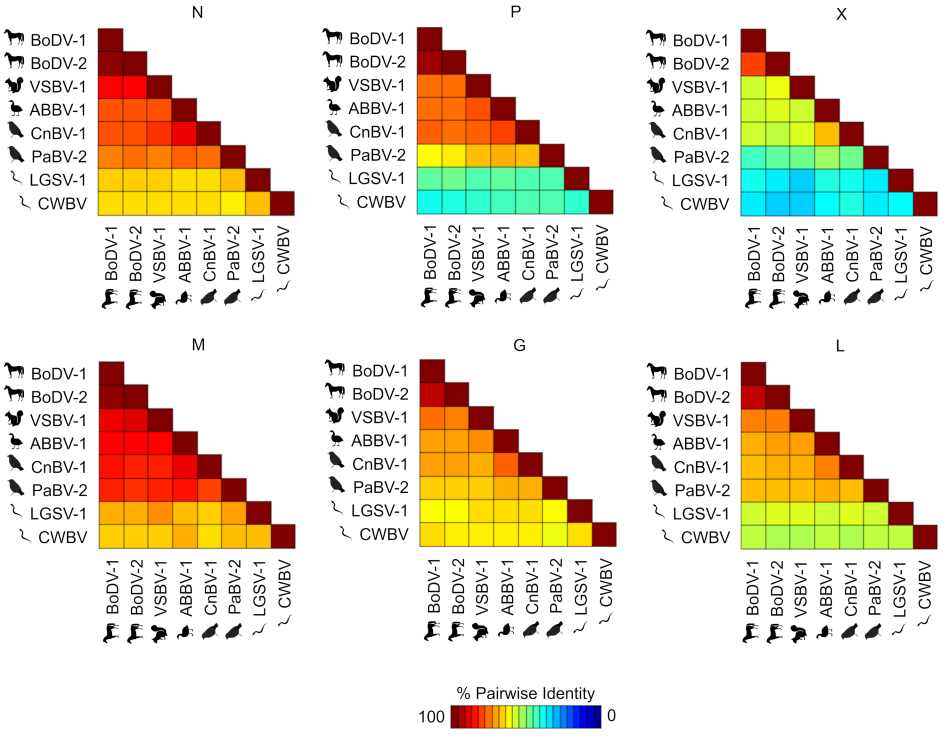
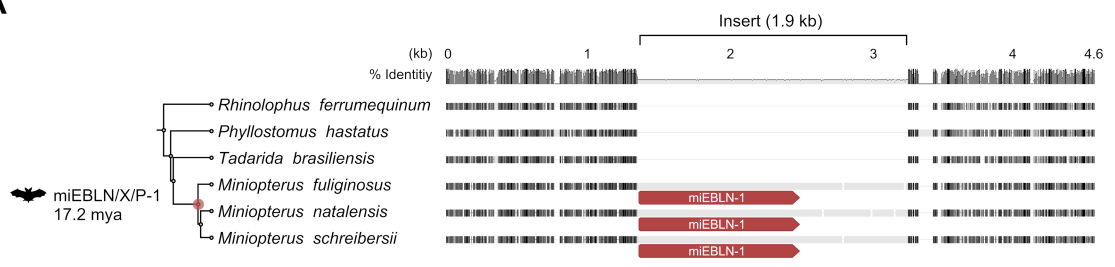
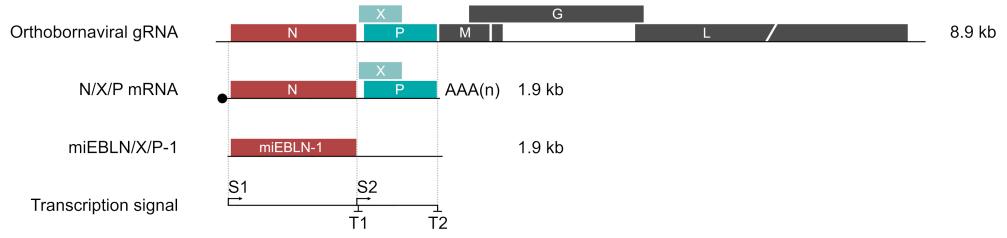

N
P
X
M
G
L
BoDV-1
BoDV-2
VSBV-1
ABBV-1
CnBV-1
PaBV-2
LGSV-1
CWBV
% Pairwise Identity
100
0

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
miEBLN-1
miEBLN-1

B


Orthobornaviral gRNA
N
X
P
M
G
L
8.9 kb
N/X/P mRNA
N
X
P
AAA(n)
1.9 kb
miEBLN/X/P-1
miEBLN-1
1.9 kb
Transcription signal
S1
S2
T1
T2

C

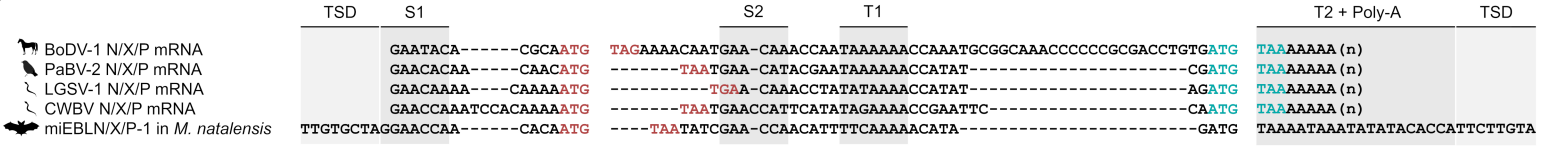

TSD
S1
S2
T1
T2 + Poly-A
TSD
BoDV-1 N/X/P mRNA
PaBV-2 N/X/P mRNA
LGSV-1 N/X/P mRNA
CWBV N/X/P mRNA
miEBLN/X/P-1 in M. natalensis
GAATACA------CGCAATG TAGAAAACAATGAA-CAAACCAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTGATG TAAAAAAA(n)
GAACACAA-----CAACATG -------TAATGAA-CATACGAATAAAAAACCATAT----------CGATG TAAAAAAA(n)
GAACAAAA----CAAAATG ----------TGAA-CAAACCTATATAAAACCATAT----------AGATG TAAAAAAA(n)
GAACCAAATCCACAAAATG -------TAATGAACCATTCATATAGAAAACCGAATTC----------CAATG TAAAAAAA(n)
TTGTGCTAGGAACCAA------CACAATG ----TAATATCGAA-CCAACATTTTCAAAAACATA----------GATG TAAAATAAATATATACACCATTCTTGTA

Fig. 3

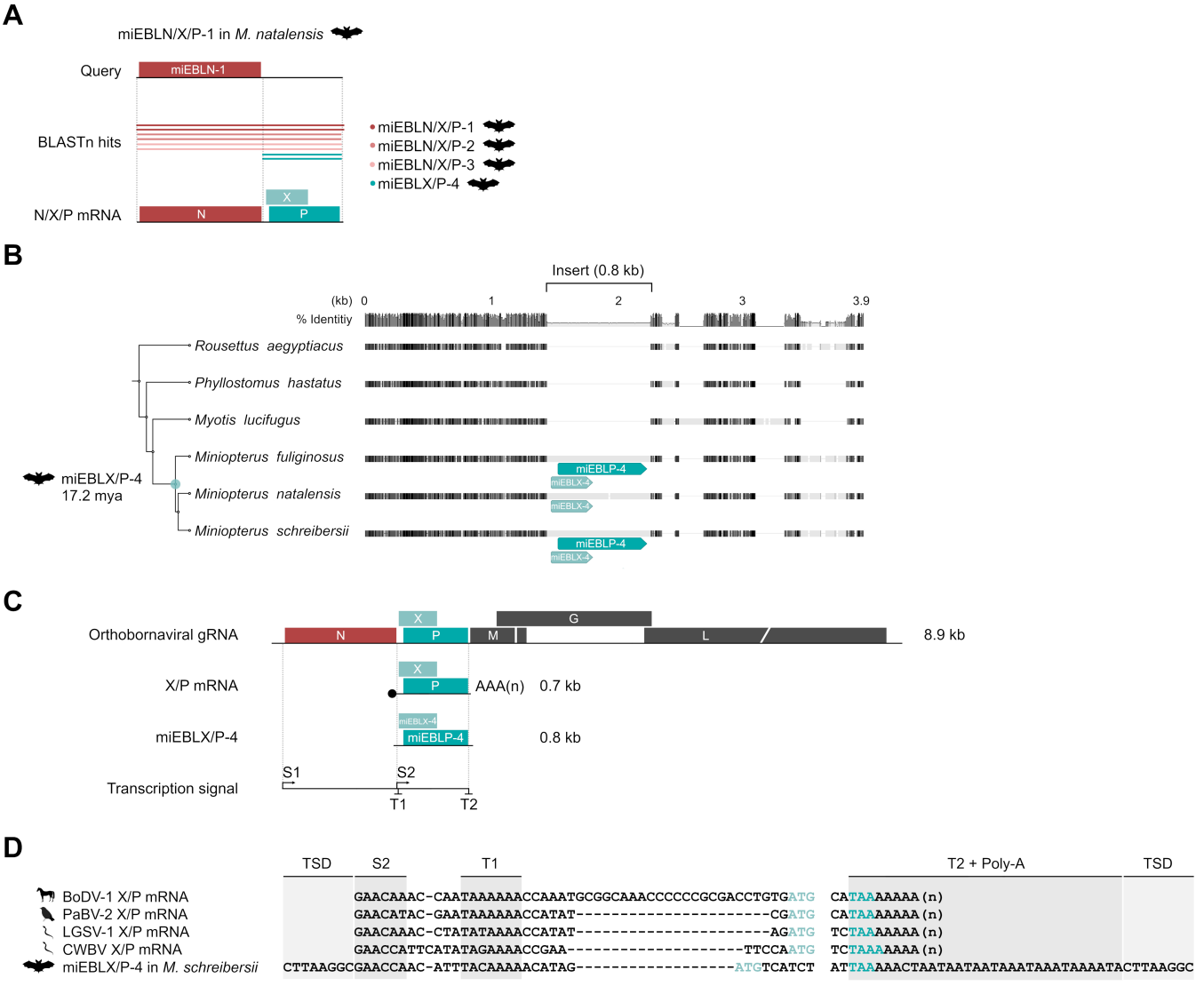

A
miEBLN/X/P-1 in M. natalensis
Query
miEBLN-1
BLASTn hits
N/X/P mRNA
N
X
P
miEBLN/X/P-1
miEBLN/X/P-2
miEBLN/X/P-3
miEBLX/P-4
B
Insert (0.8 kb)
(kb) 0
1
2
3
3.9
% Identitiy
Rousettus aegyptiacus
Phyllostomus hastatus
Myotis lucifugus
Miniopterus fuliginosus
Miniopterus natalensis
Miniopterus schreibersii
miEBLX/P-4
17.2 mya
miEBLP-4
miEBLX-4
C
Orthobornaviral gRNA
N
X
P
M
G
L
8.9 kb
X/P mRNA
AAA(n)
0.7 kb
miEBLX/P-4
miEBLX-4
miEBLP-4
0.8 kb
Transcription signal
S1
S2
T1
T2
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

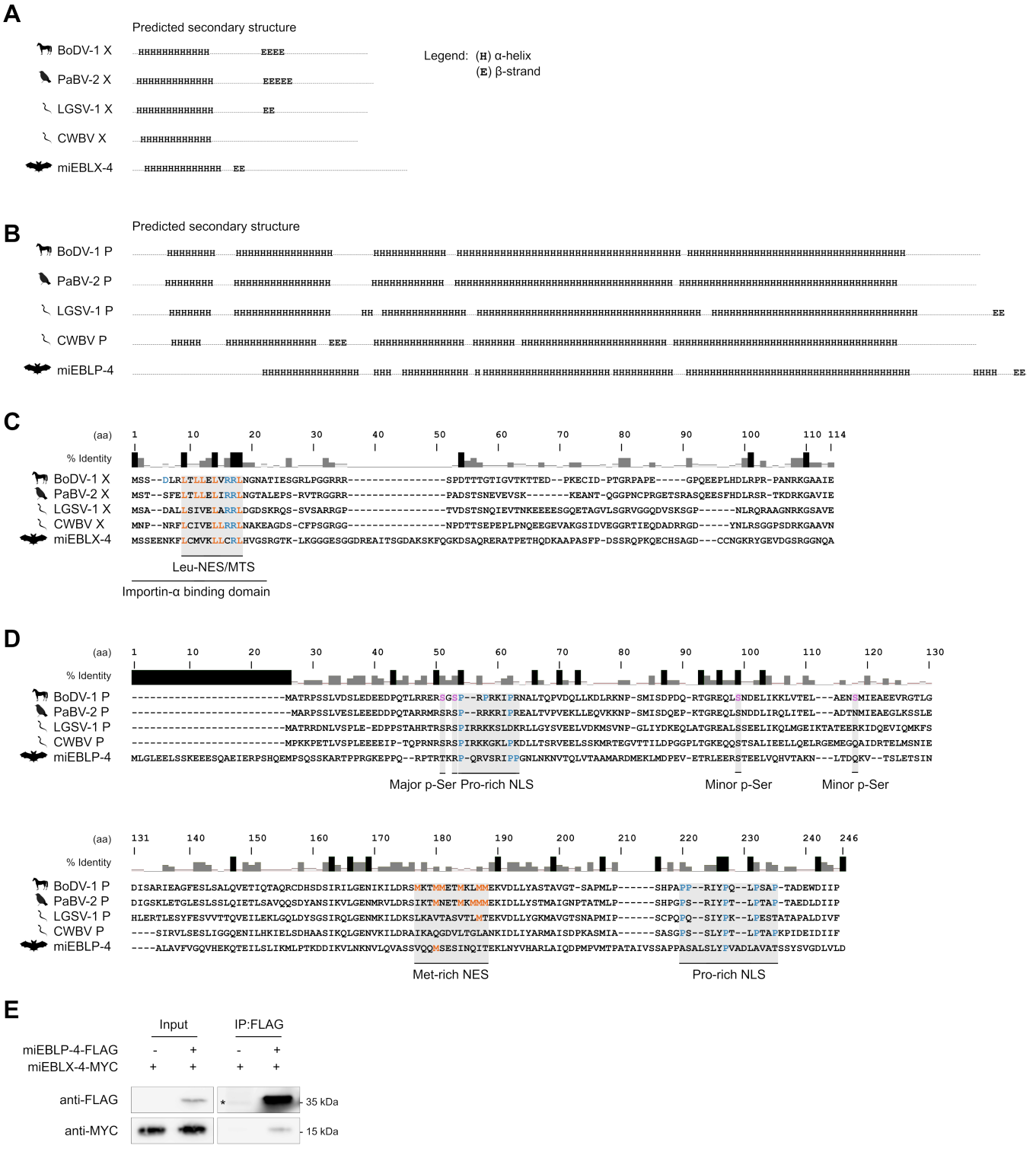
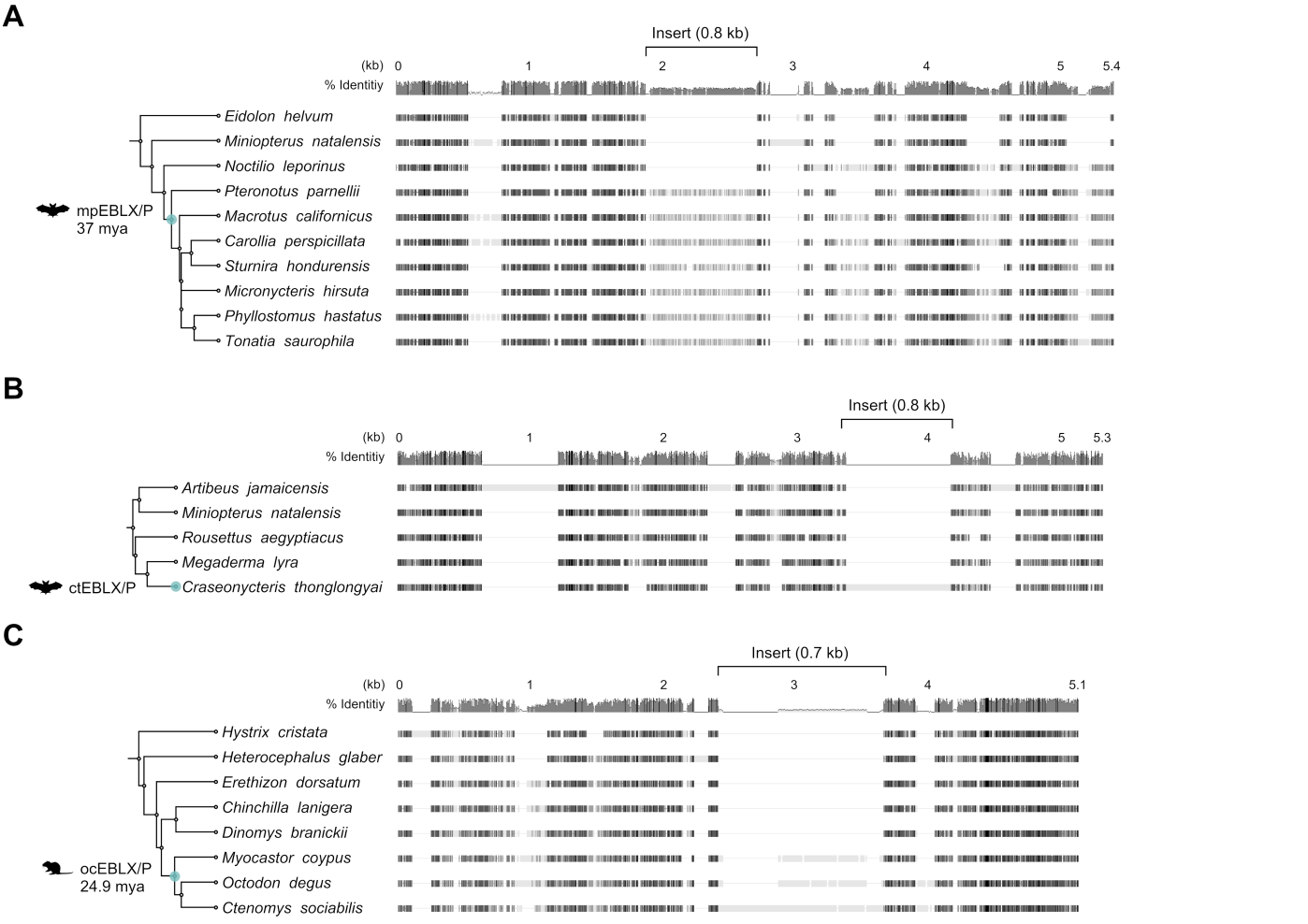

A
Predicted secondary structure
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX-4
Legend: (H) α-helix
(E) β-strand
B
Predicted secondary structure
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP-4
C
(aa)
% Identity
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX-4
Leu-NES/MTS
Importin-α binding domain
D
(aa)
% Identity
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP-4
Major p-Ser Pro-rich NLS
Minor p-Ser
Minor 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

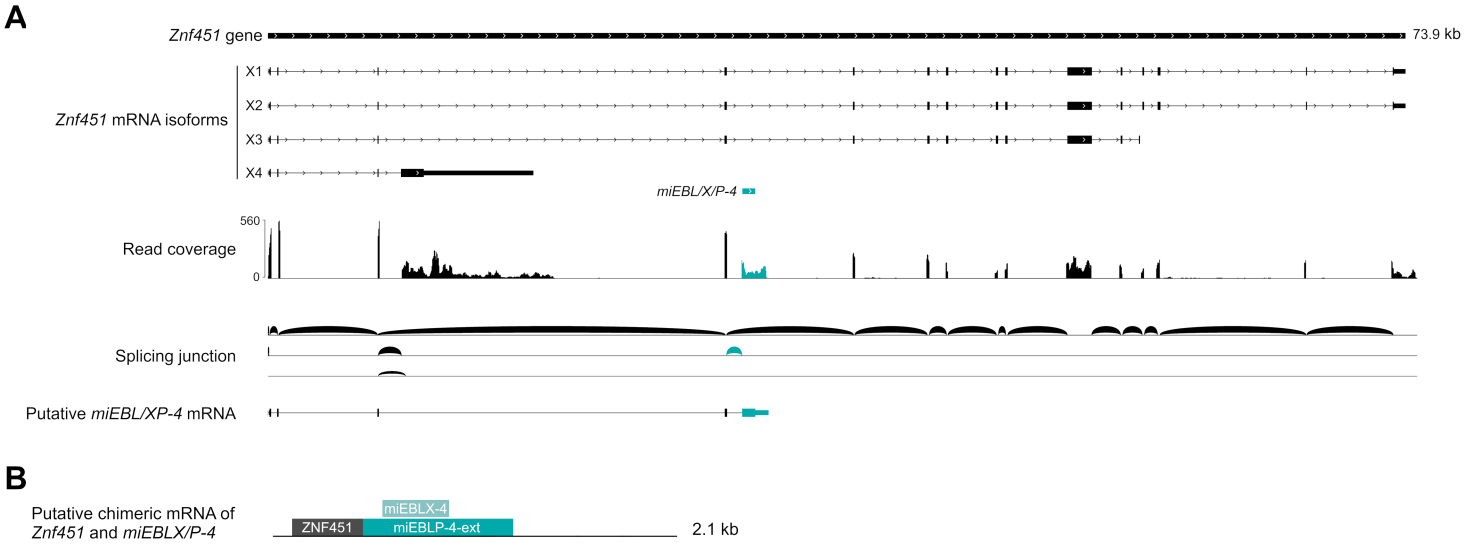
Fig. 5



D

	TSD	S2	T1		T2 + Poly-A	TSD
BoDV-1 X/P mRNA		GAACAAAC-	CAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTG	ATG	CATAAAAAAA(n)	
PaBV-2 X/P mRNA		GAACATAC-	GAATAAAAAACCATAT----------	CG	ATG	CATAAAAAAA(n)	
LGSV-1 X/P mRNA		GAACAAAC-	CTATATAAAACCATAT----------	AG	ATG	TCTAAAAAAA(n)	
CWBV X/P mRNA		GAACCATTCATATAGAAAACCGAA-	----------	TTCCA	ATG	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
miEBL/X/P-4
Read coverage
560
0
Splicing junction
Putative miEBL/XP-4 mRNA
B
Putative chimeric mRNA of
Znf451 and miEBLX/P-4
miEBLX-4
ZNF451
miEBLP-4-ext
2.1 kb