Unraveling the hidden diversity of endogenous bornavirus-like elements derived from ancient orthobornaviral X and P genes in mammalian 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, which are strongly suggested to originate from transcripts of ancient bornaviruses. EBLs have been detected by sequence similarity searches, such as tBLASTn, and thus EBLs derived from small and/or rapidly evolving viral genes, such as viral X and P genes, are difficult to be detected. Indeed, no EBLs derived from X and P genes of orthobornaviruses were detected. Thus, although previous studies comprehensively analyzed the presence of EBLs in vertebrate genomes, there may still be undetectable EBLs. In this study, we developed a novel strategy to detect such "hidden" EBLs by focusing on the nature of readthrough transcription of orthobornavirus. We showed a series of evidence supporting that there are EBLs derived from orthobornaviral X and P genes in mammalian genomes. Therefore, this study contributes to a deeper understanding of the evolution of ancient viruses and of virus-host relationships. Further, our data also suggest that endogenous viral elements detected thus far are just the tip of the iceberg, and thus further studies are required to understand ancient viruses more accurately.

Introduction

Ancient viruses from millions of years ago left fossils buried within the genomes of organisms in the form of endogenous viral elements (EVEs) [1]. These unconventional molecular fossils were derived via integration of viral sequences in the germline DNA of their hosts and subsequential transmission of the integrated sequences to the host offspring [1]. As fossils of ancient viruses, EVEs have facilitated the study of diversity and long-term evolution of viruses. Further, some EVEs were eventually co-opted by the hosts to function as novel gene regulatory elements, non-coding RNAs, and proteins [2].

Most of the discovered EVEs are derived from retroviruses due to reverse transcription and integration of viral genomes into the host genomes as part of the retroviral infection cycle using the virally encoded enzymes. However, non-retroviral EVEs have also been discovered. Endogenous bornavirus-like elements (EBLs) are EVEs that are derived from bornavriuses (family *Bornaviridae*), a group of non-segmented negative strand RNA viruses [3]. The family *Bornaviridae* is further divided into three genera: *Orthobornavirus*, *Carbovirus*, and *Cultervirus* [4]. Our previous study revealed that almost all the detected EBLs are divided into three clades corresponding to the classification of modern bornaviruses [5]. A series of evidence strongly suggest that EBLs have been formed by reverse-transcription and integration of viral mRNA via the activity of LINE-1, a retrotransposon [3, 6]. Bornaviruses have six genes, N, X, P, M, G, and L, among which N, P, M, G, and L-derived EBLs (EBLN, EBLP, EBLM, EBLG, and EBLL) were identified thus far [5].

There is a bias on the numbers of detected integrated viral genes (transcripts): among the identified EBLs, no EBLs derived from X or P genes were found, with the exception of a few carboviral EBLPs [5, 7]. There are two hypotheses to explain the bias. First, X/P

transcripts may be rarely reverse-transcribed and/or integrated into the host genome. However, the amount of X/P transcript is comparable to that of the N transcript in cells infected with Borna disease virus 1 (BoDV-1), a modern orthobornavirus [8], suggesting that there is enough chance to be integrated into the host genome. Further, we experimentally demonstrated that X/P mRNA of BoDV-1 could be integrated into the host genomes [9]. The second hypothesis is that EBLX/Ps are present, but we may not be able to detect them by the conventional sequence similarity searches due to the technical limitation. EBLs were mainly detected by tBLASTn-based method, therefore, small and rapidly evolving genes are difficult to be detected [7]. Indeed, X and P genes are small and the least conserved among the genes encoded by bornaviruses (Fig. 1). If we could detect such hidden EBLs, it would contribute to a deeper understanding of ancient viruses and virus-host coevolution.

In this study, we developed a novel strategy to detect EBLs that were previously undetectable by focusing on the nature of readthrough transcription of orthobornaviruses. We here showed a series of evidence supporting that there exist several EBLX/Ps that were undetectable in the conventional tBLASTn searches.

Results

Identification of N/X/P transcript-like EBLs in miniopterid bat genomes

We previously found miEBLN-1, which is derived from an ancient orthobornaviral N gene and retains the intact ORF, in miniopterid bats (Fig. 2a) and the miEBLN-1 "empty" loci in the genomes of *Rhinolophus ferrumequinum* and *Phyllostomus hastatus* [10]. To confirm these results, we reanalyzed the miEBLN-1 empty loci using the current genome datasets. We confirmed the above results and further identified an miEBLN-1 empty locus

in *Tadarida brasiliensis* (Fig. 2c and Fig. S1a). Interestingly, the insertion is 1.9 kb consisting of 1.2-kb miEBLN-1 region and 0.7-kb unannotated region.

To obtain insight into the origin of the 0.7-kb region, we performed a BLASTn search against NCBI nt database using the sequence of 0.7-kb region as a query. We did not find any sequence hit except for the miEBLN-1 mRNA sequences (data not shown). We and others previously found a chimeric EBL integrant consisting of N mRNA and a retrotransposon, short interspersed nuclear element (SINE) [3, 6]. To assess whether the 0.7-kb unannotated region is SINE or not, we performed a BLASTn search against the miniopterid bat genomes using the nucleotide sequence of 0.7-kb region as a query, showing that there are only 5 BLAST hits in each *M. natalensis* and *M. schreibersii* genomes that are considered to be other EBLs (see below sections). We also analyzed the sequence by CENSOR [11], but no hit was detected in the 0.7-kb region. These results suggest that the 1.9 kb insertion is not derived from a chimeric integrant of N mRNA and SINE.

The 1.9-kb insertion is almost identical in length to the N/X/P mRNA of orthobornaviruses (Fig. 2a), and thus we speculated that this insertion is derived from an ancient N/X/P mRNA. To assess the above hypothesis, we analyzed the sequences of the 1.9-kb insertion and its flanking region. We found viral transcription signal-like sequences (S1, T1/S2, and T2) at reasonable positions and a poly-A stretch downstream of T2 (Fig. 2b and Fig. S2). This sequence organization is almost identical to that of the N/X/P mRNA of orthobornaviruses (Fig. 2a). Thus, these results suggest that the 1.9-kb insertion including miEBLN-1 is derived from an ancient N/X/P transcript. Thus, we tentatively named this insertion miEBLN/X/P-1.

Detection of additional N/X/P transcript-like EBLs in miniopterid bat genomes

Given that miEBLN/X/P-1 is derived from an N/X/P mRNA, other N/X/P and X/P mRNA-derived EBLs may be present in bat or even other animal genomes. To assess this possibility, we performed a BLASTn search using EBLN/X/P-1 as a query against bat (order Chiroptera) genomes in the NCBI WGS database. We obtained 8 BLAST hits, which can be aligned to the whole N/X/P or X/P regions in miEBLN/X/P-1, from each of the *M. natalensis* and *M. schreibersii* genomes (Fig. S3 and Table S1).

The two of three BLAST hits from each of the miniopterid bat genomes are alignable to the entire region of miEBLN/X/P-1 and show high sequence similarities (76-84% nucleotide identities) to miEBLN/X/P-1 (Table S1). The flanking region of each hit can be aligned between *M. natalensis* and *M. schreibersii*, implying that they are orthologous (Fig. 2d and Fig. S1b). Therefore, we tentatively named these BLAST hits miEBLN/X/P-2 and miEBLN/X/P-3 (Table S1). We also analyzed the genome of *M. fuliginosus* by PCR and Sangar-sequencing, showing that *M. fuliginosus* also have these EBLs (Fig. S1).

Both miEBLN/X/P-2 and miEBLN/X/P-3 retain viral transcription signal-like sequences, S1, T1/S2, and T2 (Fig. 2b and Fig. S2). Additionally, both has a poly-A stretch downstream of the T2-like sequence. Further a tandem duplication (TTGTGCTA/TGTTCA) and (GAGGAACG/GAGGGATG) was also observed immediately upstream and downstream of S1 and T2-poly-A sequences of miEBLN/X/P-2 and miEBLN/X/P-3, respectively, which is probably a hallmark of LINE-1-mediated integration, called target site duplication (TSD). These results strongly suggest that miEBLN/X/P-1, 2, and 3 are derived from ancient N/X/P mRNAs that were processed by LINE-1.

138 ***Detection of X/P transcript-like EBLs in miniopterid bat genomes***

139 By the above BLASTn analysis, we also found approximately 0.8-kb BLAST hits that
140 can be aligned only to the putative X/P region (0.7-kb region) of miEBLN/X/P-1 with
141 75% nucleotide identity (Fig. S3 and Table S1). These BLAST hits are orthologous
142 between *M. natalensis* and M. *schreibersii* as their flanking regions were well aligned
143 (Fig. 3c and Fig. S4a). We found that viral transcription signal-like sequences, S2/T1 and
144 T2, are located at the reasonable positions, and the T2-like sequences are followed by a
145 poly-A stretch (Fig. 3a-3b and Fig. S5). Further, a TSD-like sequence is surrounding the
146 BLAST hits. Further, our comparative genomic analysis identified empty loci in other
147 species of bats, revealing the integrant is approximately 0.8 kb (Fig. 3c and Fig. S4a).
148 These observations suggest that miEBLX/P-4 is derived from X/P mRNA of an
149 orthobornavirus.

150 We also found putatively intact X- and P-like ORFs in the miEBLX/P-4 of *M.*
151 *schreibersii* and *M. fuliginosus* (Fig. 3c and Fig. S4a), although the P-like ORF in *M.*
152 *natalensis* was disrupted by a premature stop codon. We therefore name these X- and P-
153 like ORFs as miEBLX and miEBLP, respectively. The miEBLX and miEBLP consist of
154 109 and 233 codons, respectively, which are slightly longer than those of
155 orthobornaviruses (X: 86-90 codons; P: 201-208 codons). The putative amino acid
156 sequences of the miEBLX and miEBLP are totally different from those of the X and P
157 proteins of orthobornaviruses. However, a series of *in silico* analyses revealed that the
158 biochemical properties of miEBLX and miEBLP are similar to those of modern
159 orthobornavirus X and P protein, respectively (Fig. 4a-4b and Fig. S6-S7).

160 Secondary structure prediction analysis showed that orthobornaviral X and miEBLX
161 have an alpha helix at the N-terminal region (Fig. 4a and Fig. S6a). Hydropathy plot

showed similar patterns of viral X and miEBLX (Fig. S7a). The detailed sequence analyses revealed that the N-terminal region is the most conserved region among the viral X and miEBLX, which houses two functional domains that were previously characterized in BoDV-1 X (Fig. 4c). The N terminal region of BoDV-1 X protein contain leucine-rich domain functioning both as a CRM1-dependent nuclear export signal and mitochondria-targeting signal [12-14].

The miEBLP protein also show similar results. A predicted secondary structure and hydropathy plot of miEBLP protein are similar to those of orthobornavirus P proteins (Fig. 4b and Fig. S6b). Hydropathy plot showed similar patterns of viral P and miEBLP (Fig. S7b). We then aligned miEBLP with the extant P proteins and checked the presence of functional domains that were previously characterized in BoDV-1 P (Fig. 4d). These results suggest that the miEBLX and miEBLP are intact ORFs.

Together with the above results, our data strongly suggest that miEBLN/X/P-1, 2, 3 and miEBLX/P-4 are derived from N/X/P and X/P transcripts of ancient orthobornaviruses.

Detection of diverse EBLX/Ps in bat and rodent genomes

To further identify EBLX/Ps in mammalian genomes, we performed a TFASTY search [15] against mammalian genomes using the putative amino acid sequences of miEBLX and miEBLP. TFASTY is a sequence similarity search program that allows frame shifts, which may be more sensitive than tBLASTn for EVE search because EVEs often contain frame shifts due to accumulation of mutations. By the TFASTY search we identified additional EBLX/Ps in bat, rodent, and solenodon genomes (Fig. S3b-S3c and Table S2-S3).

By comparative genomics, we confirmed that at least some of the detected EBL candidates are inserted sequences (Fig. 3d-3f and Fig. S4b-S4d). Further, we identified the hallmarks of EBLX/Ps in some of the TFASTY hits, such as transcription signal-like sequences and TSDs (Fig. 3b and Fig. S5). These results strongly suggest that the sequences detected by TFASTY, if not all, are indeed EBLX/Ps. Taken together, divergent EBLX/Ps are present in mammalian genomes that were not detectable by the conventional detection method using tBLASTn alone.

Discussion

Although many EBLs were identified thus far, no X and only a few P-derived EBLs were reported. There were two hypotheses to explain this observation: (i) X/P mRNA is rarely reverse-transcribed and integrated into host genomes, and (ii) there are X/P-derived EBLs which are undetectable by the conventional BLAST searches due to the technical limitation. In this study, we focused on the nature of readthrough transcription of orthobornaviruses that produces approximately 1.9-kb N/X/P transcripts and applied this to exploration of EBLs derived from viral X and P genes. We showed a series of evidence that EBLX/Ps are indeed present in mammalian genomes but not detectable by the conventional BLAST searches. Thus, this study could reveal hidden EBLX/Ps, which contribute to a deeper understanding of ancient bornaviruses.

The EBLX/Ps identified in this study may be just the tip of the iceberg: namely, we are probably still missing many EBLs derived from X/P genes. In this study, we identified the 1.9-kb integrant (miEBLN/X/P-1) in the miniopterid bat genomes by chance and did not further explore this kind of EBLs. Manual curation is time-consuming, and thus development of efficient searching method to identify approximately 1.9-kb insertion

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 was also observed in other viruses, and thus our searching strategy can be applied to detecting further EVEs.

Taken together, we identified EBLs that were undetectable using the conventional BLAST searches. Our findings provide novel insights into ancient bornaviruses. Further, our searching strategy will contribute to paleovirological research.

Materials and Methods

Sangar sequencing of EBLs and flanking regions in M. fulginosus

miEBLN/X/P-1, miEBLN/X/P-2, miEBLN/X/P-3, and miEBLX/P-4 and corresponding flanking regions in *M. fuliginosus* were PCR-amplified from DNA sample extracted from YubFKT1 (kidney epithelial cell line) [16] using Phusion Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific). The obtained amplicons were purified using innuPREP PCRpure Lite Kit (Analytik Jena), and then sequenced by Sanger dideoxy method at Eurofins Genomics. The primer sequences are available in Table S4.

BLAST searches for identification of EBLN/X/Ps and EBLX/Ps

To search for EBLX/P insertions, miEBLN/X/P-1 from *M. natalensis* was used as query for BLASTn search [17] against the WGS database [18] of bats (order Chiroptera; taxid: 9397), with changes applied to default parameters: E-value threshold = 10^{-10}, word size = 7, no filter for low complexity regions.

TFASTY search against mammalian genomes to detect EBLX/Ps

To detect EBLX/Ps in mammalian genomes, TFASTY searches were performed using the FASTA36 package [15]. Mammalian genomes were downloaded using NCBI Datasets [18], and used for the database for TFASTY search.

Identification of orthologous loci

To search for the orthologous loci harboring EBL insertions or EBL "empty" loci, the sequence of the insertion and its flanking regions were used as query for BLASTn search against the WGS contig database of Chiroptera (taxid: 9397) or Rodentia (taxid: 9989), with changes applied to default parameters: E-value = 10^{-20}, word size = 11.

The BLAST hit sequences were retrieved, which were used for further analysis. Genomic sequences from different species which corresponds to the locus containing the EBL insertion were aligned by MAFFT alignment tool. The phylogenetic tree of the species were obtained from TimeTree [19]. See Table S5 for the nucleotide accessions and regions of EBL insertions and the loci containing the insertions.

Annotation of EBLs

Following alignment, TSD sequences flanking the EBL insertions were determined and promptly annotated (if present). The EBLs were then aligned by MAFFT alignment tool, wherein transcription signal sequences S2, T1, and T2 were determined and annotated, by also aligning to extant orthobornaviral X/P transcript sequences.

Complete X- and P-like ORFs from miEBLX/P-4 were also annotated as miEBLX and miEBLP, respetively. The amino acid sequences were then extracted and aligned with the amino acid sequences of extant orthobornaviral X or P proteins. Functional domains previously characterized in BoDV-1 X and P were also determined within miEBLX and

miEBLP, as well as other extant orthobornaviral X and P proteins. See Table S6 for the nucleotide accession of extant orthobornaviruses.

All sequence alignments and annotations, as well as phylogenetic tree reconstruction, were performed using Geneious version 11.1.5.

Pairwise similarity analysis

Amino acid sequences of the viral proteins of extant orthobornaviruses (Table S6) were analyzed by MAFFT alignment tool [20] using Sequence Demarcation Tool software version 1.2 [21]. 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 and miEBLP from *M. fuliginosus*, as well as extant orthobornaviral X and P proteins, were submitted to JPred version 4 [22]. The simple results were then exported for further analysis.

To determine the protein hydrophathy, the same amino acid sequences were submitted to ProtScale in ExPASy Server [23], using the amino acid scale developed by Kyte & Doolittle [24]. The window sizes were set to 9 and 15 for miEBLX and miEBLP, respectively. The hydropathy scores were then exported for further analysis.

Acknowledgements

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

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

Conflict of interest

None declared.

References

1. Aiewsakun, P. and A. Katzourakis, *Endogenous viruses: Connecting recent and ancient viral evolution.* Virology, 2015. **479-480**: p. 26-37.
2. Jern, P. and J.M. Coffin, *Effects of retroviruses on host genome function.* Annu Rev Genet, 2008. **42**: p. 709-32.
3. Horie, M., et al., *Endogenous non-retroviral RNA virus elements in mammalian genomes.* Nature, 2010. **463**(7277): p. 84-7.
4. Rubbenstroth, D., et al., *ICTV Virus Taxonomy Profile: Bornaviridae.* J Gen Virol, 2021. **102**(7).
5. Kawasaki, J., et al., *100-My history of bornavirus infections hidden in vertebrate genomes.* Proc Natl Acad Sci U S A, 2021. **118**(20).
6. Belyi, V.A., A.J. Levine, and A.M. Skalka, *Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes.* PLoS Pathog, 2010. **6**(7): p. e1001030.
7. Horie, M. and K. Tomonaga, *Paleovirology of bornaviruses: What can be learned from molecular fossils of bornaviruses.* Virus Res, 2019. **262**: p. 2-9.
8. Kojima, S., et al., *Splicing-Dependent Subcellular Targeting of Borna Disease Virus Nucleoprotein Isoforms.* J Virol, 2019. **93**(5).
9. Horie, M., et al., *Comprehensive analysis of endogenous bornavirus-like elements in eukaryote genomes.* Philos Trans R Soc Lond B Biol Sci, 2013. **368**(1626): p. 20120499.
10. Mukai, Y., et al., *An endogenous bornavirus-like nucleoprotein in miniopterid bats retains the RNA-binding properties of the original viral protein.* FEBS Lett, 2022.
11. Jurka, J., et al., *CENSOR--a program for identification and elimination of repetitive elements from DNA sequences.* Comput Chem, 1996. **20**(1): p. 119-21.
12. Ferre, C.A., 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): p. 1523-33.

13. Poenisch, M., et al., *Protein X of Borna disease virus inhibits apoptosis and promotes viral persistence in the central nervous systems of newborn-infected rats.* J Virol, 2009. **83**(9): p. 4297-307.
14. Szelechowski, M., et al., *A viral peptide that targets mitochondria protects against neuronal degeneration in models of Parkinson's disease.* Nat Commun, 2014. **5**: p. 5181.
15. Pearson, W.R., *Finding Protein and Nucleotide Similarities with FASTA.* Curr Protoc Bioinformatics, 2016. **53**: p. 3 9 1-3 9 25.
16. Maeda, K., et al., *Isolation of novel adenovirus from fruit bat (Pteropus dasymallus yayeyamae).* Emerg Infect Dis, 2008. **14**(2): p. 347-9.
17. Camacho, C., et al., *BLAST+: architecture and applications.* BMC Bioinformatics, 2009. **10**: p. 421.
18. Coordinators, N.R., *Database resources of the National Center for Biotechnology Information.* Nucleic Acids Res, 2018. **46**(D1): p. D8-D13.
19. Kumar, S., et al., *TimeTree: A Resource for Timelines, Timetrees, and Divergence Times.* Mol Biol Evol, 2017. **34**(7): p. 1812-1819.
20. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7: improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.
21. Muhire, B.M., A. Varsani, and D.P. Martin, *SDT: a virus classification tool based on pairwise sequence alignment and identity calculation.* PLoS One, 2014. **9**(9): p. e108277.
22. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015. **43**(W1): p. W389-94.
23. Duvaud, S., et al., *Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users.* Nucleic Acids Research, 2021. **49**(W1): p. W216-W227.
24. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein.* J Mol Biol, 1982. **157**(1): p. 105-32.

Figure legends

Figure 1. Amino acid identities within extant orthobornaviral proteins. Pairwise sequence identities between amino acid sequences of extant orthobornaviral proteins are shown. The accession numbers used in this analysis was shown in Table S6.

Figure 2. Characterization of EBLs derived from N/X/P transcripts. (A) A schematic figure of orthobornviral genomic RNA, its N/X/P mRNA, transcription signals and miEBLN-1 locus. S1/S2 and T1/T2 show transcription initiation and termination signals, respectively. **(B)** A nucleotide alignment of orthobornaviral N/X/P transcripts and miEBLN/X/P loci. Red letters represent the start and stop codons of viral N gene. Blue letters indicate the start codon of X gene and stop codon of P gene. TSD, target site duplication. **(C** and **D)** Gene orthology analyses for miEBLN/X/P-1 and 3. Nucleotide alignment of each miEBLN/X/P and their flanking loci is shown. The corresponding species and their phylogenetic relationships are indicated on the left. The minimum age of each EBLN/X/P is also shown.

Figure 3. Characterization of EBLs derived from X/P transcripts. (A) A schematic figure of orthobornaviral genomic RNA, its X/P mRNA, transcription signals, and miEBLX/P loci. **(B)** A nucleotide alignment of orthobornaviral X/P transcripts and miEBLX/P loci. Light blue and blue letters represent the start codon of X gene and stop codon of P gene, respectively. TSD, target site duplication. **(C - F)** Gene orthology analyses for EBLX/Ps. Nucleotide alignment of each EBLX/P and its flanking loci is shown. The corresponding species and their phylogenetic relationships are indicated on the left. The minimum age of each EBLN/X/P is also shown.

Figure 4. ***In silico*** **characterization of putative miEBLX and miEBLP proteins. (A** and **B)** Secondary structure prediction of putative EBLX and EBLP proteins. H and E represent helical and beta extended secondary structure, respectively. **(C)** Multiple amino acid sequence alignment of orthobornaviral X and putative EBLX proteins. **(D)** Multiple amino acid sequence alignment of orthobornaviral P and putative EBLP proteins.

Fig. 1

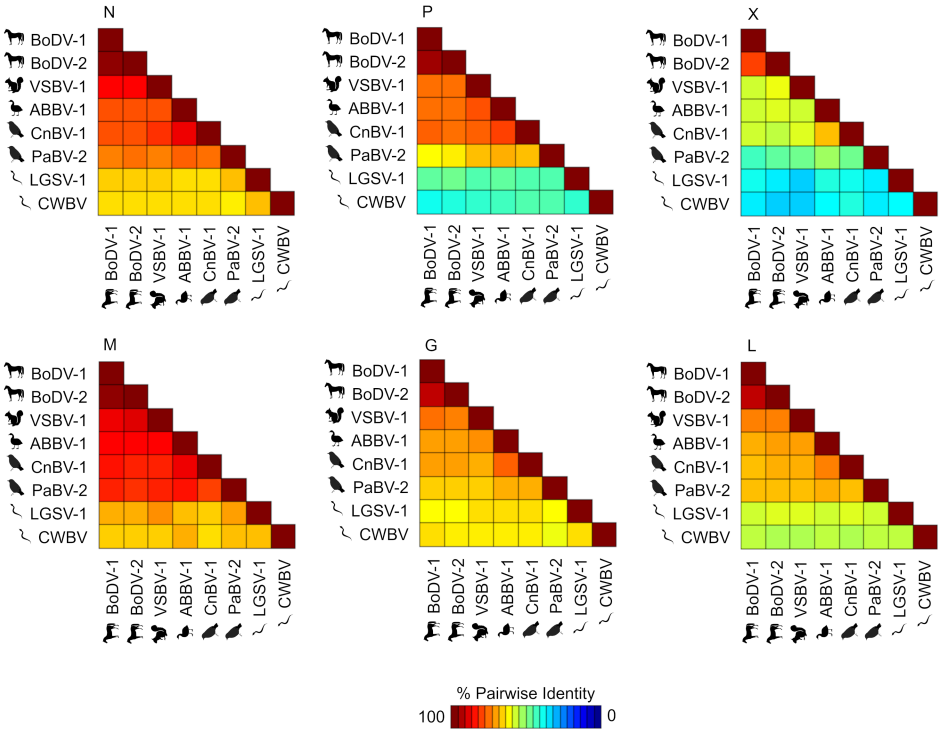
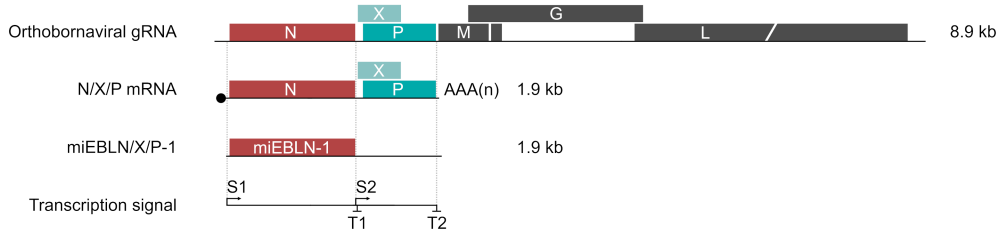
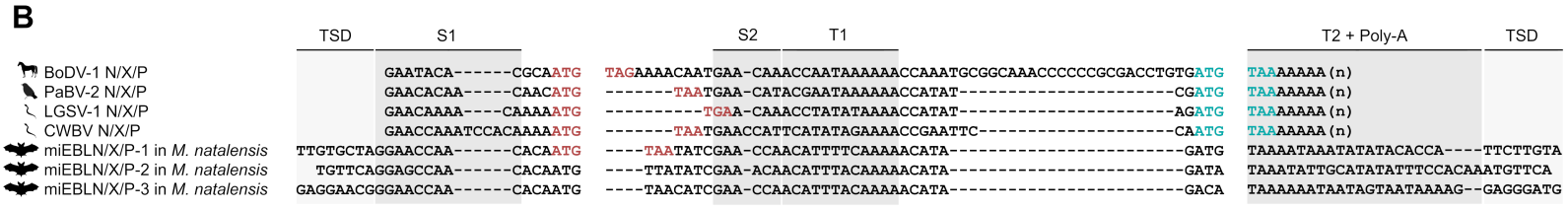

N
P
X
M
G
L
BoDV-1
BoDV-2
VSBV-1
ABBV-1
CnBV-1
PaBV-2
LGSV-1
CWBV
% Pairwise Identity
100
0

Fig. 2

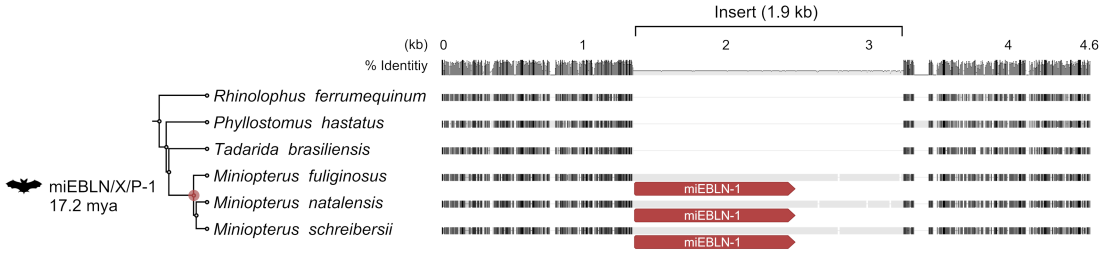
A


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

B


TSD
S1
S2
T1
T2 + Poly-A
TSD
BoDV-1 N/X/P
PaBV-2 N/X/P
LGSV-1 N/X/P
CWBV N/X/P
miEBLN/X/P-1 in M. natalensis
miEBLN/X/P-2 in M. natalensis
miEBLN/X/P-3 in M. natalensis

C


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

D

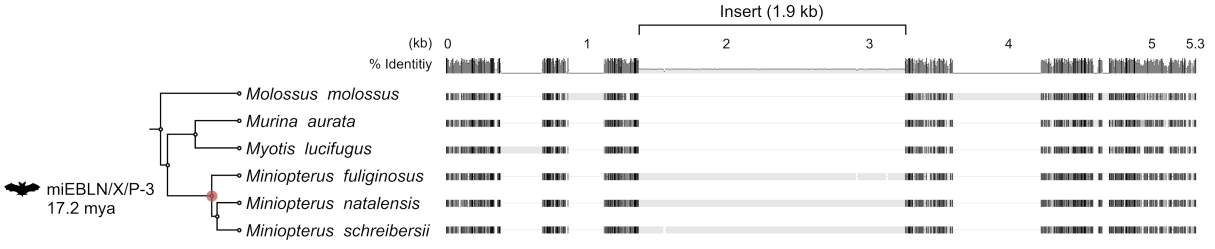

Insert (1.9 kb)
(kb)
0
1
2
3
4
5
5.3
% Identitiy
Molossus molossus
Murina aurata
Myotis lucifugus
Miniopterus fuliginosus
Miniopterus natalensis
Miniopterus schreibersii
miEBLN/X/P-3
17.2 mya

Fig. 3

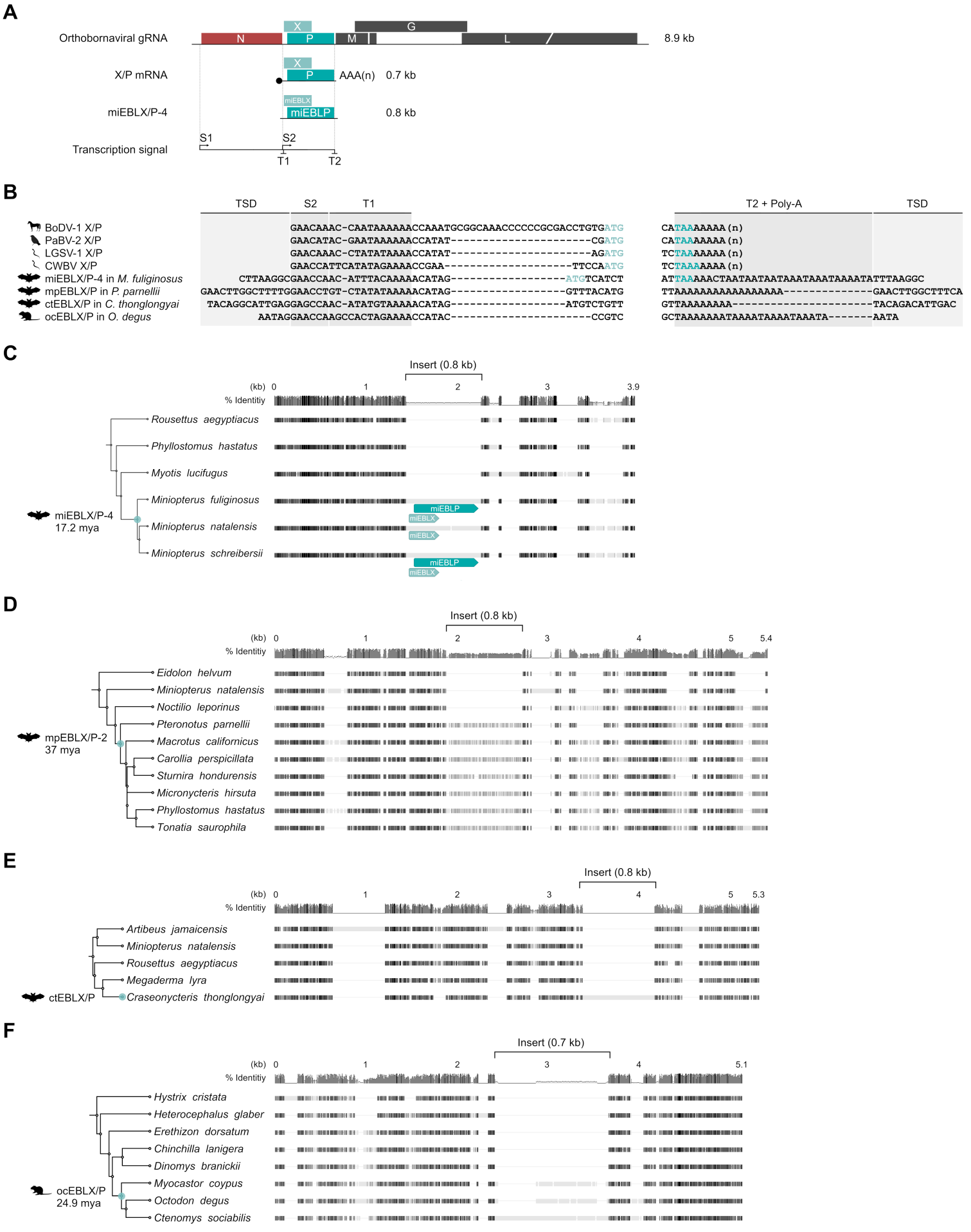
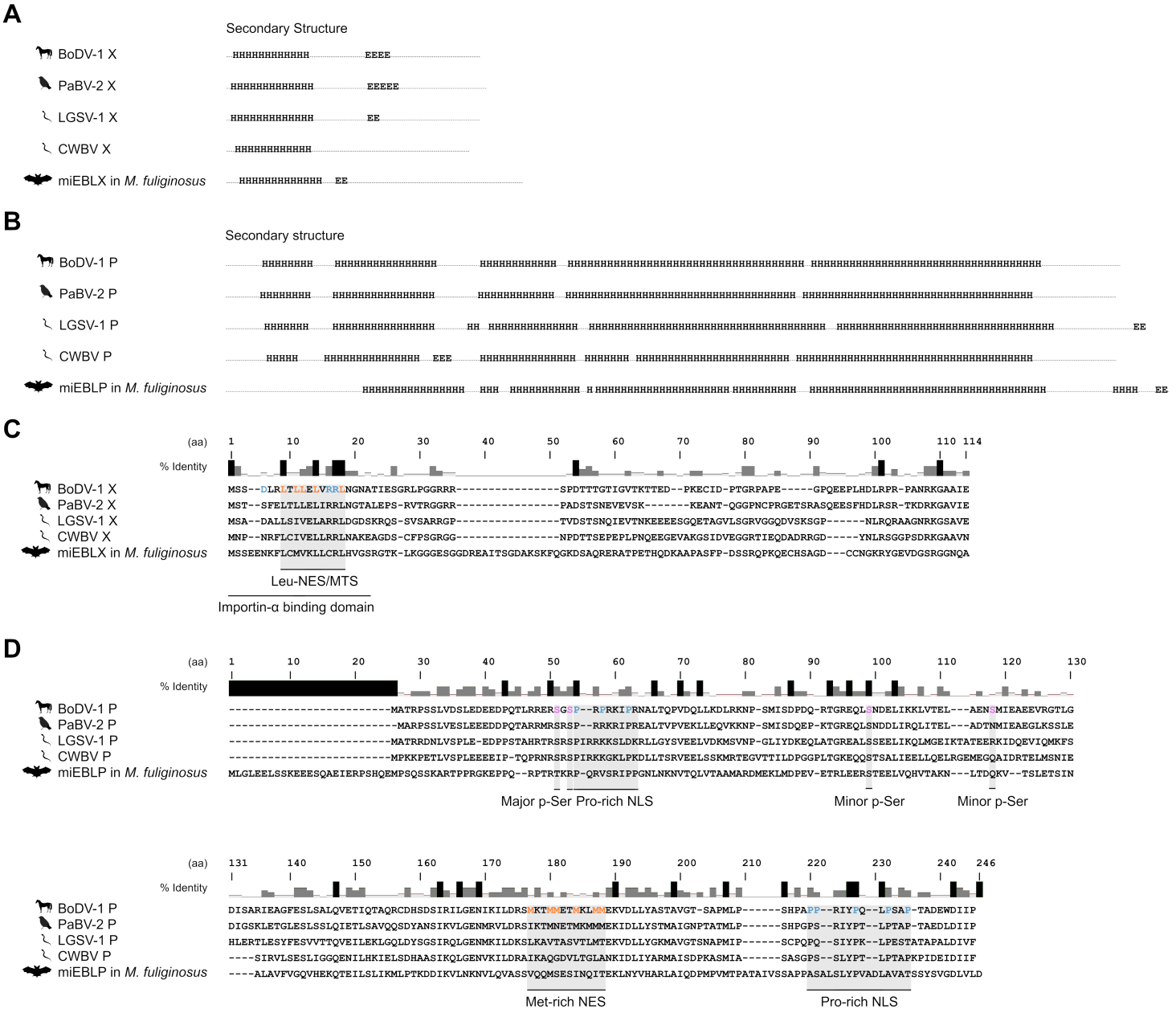

A
Orthobornaviral gRNA
N
X
P
M
G
L
8.9 kb
X/P mRNA
AAA(n)
0.7 kb
miEBLX/P-4
miEBLX
miEBLP
0.8 kb
Transcription signal
S1
S2
T1
T2
B
TSD
S2
T1
T2 + Poly-A
TSD
BoDV-1 X/P
PaBV-2 X/P
LGSV-1 X/P
CWBV X/P
miEBLX/P-4 in M. fuliginosus
mpEBLX/P in P. parnellii
ctEBLX/P in C. thonglongyai
ocEBLX/P in O. degus
GAACAAAC-CAATAAAAAACCAAATGCGGCAAACCCCCCGCGACCTGTGATG
GAACATAC-GAATAAAAAACCATAT----------CGATG
GAACAAAC-CTATATAAAACCATAT----------AGATG
GAACCATTCATATAGAAAACCGAA----------TTCCAATG
CTTAAGGCGAACCAAC-ATTTACAAAAACATAG----------ATGTCATCT
GAACTTGGCTTTTGGAACCTGT-CTATATAAAAACATAG----------GTTTACATG
TACAGGCATTGAGGAGCCAAC-ATATGTAAAAACATAG----------ATGTCTGTT
AATAGGAACCAAGCCACTAGAAAACCATAC----------CCGTC
CATAAAAAAA(n)
CATAAAAAAA(n)
TCTAAAAAAA(n)
TCTAAAAAAA(n)
ATTAAAACTAATAATAATAAATAAATAAAATATTTAAGGC
TTAAAAAAAAAAAAAAAAA----------GAACTTGGCTTTCA
GTTAAAAAAAA----------TACAGACATTGAC
GCTAAAAAAATAAAATAAAATAAATA-------AATA
C
Insert (0.8 kb)
(kb) 0 1 2 3 3.9
% Identitiy
Rousettus aegyptiacus
Phyllostomus hastatus
Myotis lucifugus
Miniopterus fuliginosus
Miniopterus natalensis
Miniopterus schreibersii
miEBLP
miEBLX
miEBLX/P-4
17.2 mya
D
Insert (0.8 kb)
(kb) 0 1 2 3 4 5 5.4
% Identitiy
Eidolon helvum
Miniopterus natalensis
Noctilio leporinus
Pteronotus parnellii
Macrotus californicus
Carollia perspicillata
Sturnira hondurensis
Micronycteris hirsuta
Phyllostomus hastatus
Tonatia saurophila
mpEBLX/P-2
37 mya
E
Insert (0.8 kb)
(kb) 0 1 2 3 4 5 5.3
% Identitiy
Artibeus jamaicensis
Miniopterus natalensis
Rousettus aegyptiacus
Megaderma lyra
Craseonycteris thonglongyai
ctEBLX/P
F
Insert (0.7 kb)
(kb) 0 1 2 3 4 5.1
% Identitiy
Hystrix cristata
Heterocephalus glaber
Erethizon dorsatum
Chinchilla lanigera
Dinomys branickii
Myocastor coypus
Octodon degus
Ctenomys sociabilis
ocEBLX/P
24.9 mya

Fig. 4


A
Secondary Structure
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX in M. fuliginosus
B
Secondary structure
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP in M. fuliginosus
C
(aa)
% Identity
BoDV-1 X
PaBV-2 X
LGSV-1 X
CWBV X
miEBLX in M. fuliginosus
Leu-NES/MTS
Importin-α binding domain
D
(aa)
% Identity
BoDV-1 P
PaBV-2 P
LGSV-1 P
CWBV P
miEBLP in M. fuliginosus
Major p-Ser Pro-rich NLS
Minor p-Ser
Minor p-Ser
Met-rich NES
Pro-rich NLS