Virus Evolution, 2026, **12(1)**, veag033
https://doi.org/10.1093/ve/veag033
Advance Access Publication 5 June 2026
Research Article

Continuous *in silico* screening of fish, amphibians, reptiles, and birds expands the diversity of bornavirids and reveals unexpected genomic architectures

Mirette I. Y. Eshak[1,†], Sakiho Imai[2,†], Junna Kawasaki[3], Mai Kishimoto[2,4], Miguel Vences[5], Andolalao Rakotoarison[6,7], Martin Beer[1], Dennis Rubbenstroth[1], Masayuki Horie[2,4,*], Florian Pfaff[1,*]

[1]Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, 17493 Greifswald—Insel Riems, Germany
[2]Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, 1-58 Rinku Orai-kita, Izumisano 598-8531, Osaka, Japan
[3]Department of Infectious Disease Pathobiology, Graduate School of Medicine, Chiba University, 1-8-1, Inohana, Chuo-ku, Chiba-shi, Chiba 260-8675, Japan
[4]Osaka International Research Center for Infectious Diseases, Osaka Metropolitan University, 32-1 Morinomiya 2-chome, Joto-ku, Osaka 536-8525, Japan
[5]Zoological Institute, Technische Universität Braunschweig, Mendelssohnstraße 4, 38106 Braunschweig, Germany
[6]Mention Environnement, Université de l'Itasy, Faliarivo Ambohidanerana, 118 Soavinandriana Itasy, Madagascar
[7]School for International Training, VN 41A Bis Ankazolava Ambohitsoa, Antananarivo 101, Madagascar
*Corresponding authors. Masayuki Horie, Laboratory of Veterinary Microbiology, Graduate School of Veterinary Science, Osaka Metropolitan University, 1-58 Rinku Orai-kita, Izumisano 598-8531, Osaka, Japan. E-mail: mhorie@omu.ac.jp; Florian Pfaff, Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, 17493 Greifswald—Insel Riems, Germany. E-mail: florian.pfaff@fli.de.
[†]Mirette I. Y. Eshak and Sakiho Imai equal contribution

Abstract

Using an *in silico* data mining approach, we previously expanded the known diversity within the family *Bornaviridae* by identifying novel bornavirids in publicly available sequence datasets from bony and cartilaginous fish. Building on this discovery, we systematically screened numerous additional sequence datasets from fish, amphibians, reptiles, and birds in the Sequence Read Archive and a new RNA-seq dataset to further explore bornaviral diversity and evolutionary history. Our analysis identified nearly complete and partial bornaviral genomes across multiple host taxa, that represented either new viral variants, previously known viruses detected in so far unreported hosts/sources, or novel viruses. Several novel orthobornaviral genomes were identified in snake and lizard datasets. Sequences of the species *Orthobornavirus alphapsittaciforme* and *Orthobornavirus serini* were identified in avian datasets. We also found sequences of both known and possibly new species in the genus *Cultervirus*, which were detected in bony fish and, for the first time, in reptile and amphibian datasets. The identification of culterviruses in reptiles and amphibians suggests that this group of viruses, which was previously only known to be associated with fish, has a much wider host range. A genetically divergent bornaviral genome was discovered in the gold-striped salamander (*Chioglossa lusitanica*), possessing the genomic architecture of orthobornaviruses but likely representing a new genus within the family *Bornaviridae*. Interestingly, multiple distinct viral genomes identified in passerine birds were classified as carboviruses based on sequence similarities, but they exhibited distinct genomic features and lacked the genes that code for glycoprotein (G) and, in some cases, the accessory protein (X) and/or matrix protein (M). Subsequent screening of samples from passerine birds further confirmed these findings and showed that these carboviral sequences likely represent genuine viruses rather than endogenous bornaviral elements. Interestingly, these carboviruses appear to co-occur with enveloped viruses in potential co-infection or helper virus scenarios. These findings contribute substantially to our understanding of *Bornaviridae* diversity, their evolutionary adaptation as well as modes of transcriptional regulation.

Keywords virus discovery, SRA, repository, *Mononegavirales, Bornaviridae,* bornaviruses, co-infection

Introduction

The *Bornaviridae* family within the order *Mononegavirales*, currently consists of four genera: *Orthobornavirus*, *Carbovirus*, *Cultervirus*, and *Cartilovirus* (Eshak et al. 2023, Hyndman et al. 2018, Kuhn et al. 2015, Rubbenstroth et al. 2021). Members of the genus *Orthobornavirus* exhibit a remarkably broad host range, infecting avian, reptilian, and mammalian species (Rubbenstroth et al. 2021). Notably, some orthobornaviruses are known to be zoonotic, including Borna disease

Received: December 18, 2025. **Revised:** April 29, 2026. **Accepted:** May 26, 2026

virus 1 (BoDV-1; species *Orthobornavirus bornaense*) and variegated squirrel bornavirus 1 (VSBV-1; *Orthobornavirus sciuri*), both of which have been implicated in fatal encephalitis in humans (Hoffmann et al. 2015, Niller et al. 2020). In contrast, carboviruses and culterviruses have thus far been identified exclusively in reptiles or fish, respectively, with no documented cases of zoonotic transmission. The genus *Cartilovirus* is currently represented by little skate bornavirus (LSBV, *Cartilovirus plani*; GenBank accession: BK063518) (Eshak et al. 2023).

Bornavirids are enveloped viruses containing a linear, negative-sense, single-stranded RNA genome of approximately 9000 nucleotides (nt) (Briese et al. 1994, Rubbenstroth et al. 2021). Their genomes encode at least six open reading frames (ORFs), for the primary viral proteins nucleoprotein (N), accessory protein (X), phosphoprotein (P), matrix protein (M), glycoprotein (G), and the large protein (L), which functions as an RNA-dependent RNA polymerase (RdRp). The ORFs are typically arranged as 3′-N-X/P-M-G-L-5′ in orthobornaviruses and 3′-N-X/P-G-M-L-5′ in carboviruses and culterviruses (Briese et al. 1994, Rubbenstroth et al. 2021). LSBV possesses the largest bornavirid genome identified to date, which comprises 11 090 nt. This genome features a unique organization, incorporating two additional putative ORFs encoding viral proteins 1 and 2 (vp1 and 2), resulting in a third type of genomic arrangement 3′-N-vp1-vp2-X/P-G-M-L-5′ (Eshak et al. 2023). The replication and transcription of bornavirids occur within the host cell nucleus, as demonstrated in studies of BoDV-1 (Briese et al. 1992). The transcriptional complexity of orthobornaviruses and culterviruses is driven by a combination of regulated transcription initiation and termination sites (Eshak et al. 2025, Eshak et al. 2023, Schneemann et al. 1994), along with alternative splicing mechanisms that process polycistronic primary transcripts to generate a diverse array of mRNAs (Eshak et al. 2025, Eshak et al. 2023, Schneemann et al. 1994, Schneider et al. 1994, Tomonaga et al. 2000).

With advancements of computational biology, data mining has become a powerful tool for discovering novel viruses, including bornavirids. Our latest *in silico* screening expanded the genus *Cultervirus* with four novel species, along with a new variant (BK063520) of Wǔhàn sharpbelly bornavirus (WhSBV, *Cultervirus hemicultri;* NC_055169), detected in CIK and L8824 cells lines derived from grass carp (*Ctenopharyngodon idella*). In addition, we reported the first complete genome of the Murray–Darling carp bornavirus (MDCBV, *Cultervirus hemicultri*; BK063521) from a goldfish (*Carassius auratus*). Furthermore, the discovery of LSBV led to the recent proposal of the genus *Cartilovirus* (Eshak et al. 2023).

In this study, we aimed to further expand the known diversity of the *Bornaviridae* family by analysing 184 749 publicly available RNA sequencing (RNA-seq) datasets from fish, amphibians, reptiles, and birds. Our *in silico* screening of the Sequence Read Archive (SRA) identified coding-complete bornaviral genomes, representing new variants, known viruses detected in previously unreported hosts/sources, or novel viruses. Interestingly, viral genomes were identified in avian (order Passeriformes) datasets, that phylogenetically clustered with the genus *Carbovirus* but showed a unique genome organization, lacking G and some cases also X and/or M genes. One of these novel carboviruses was subsequently confirmed by reverse transcription quantitative polymerase chain reaction (RT-qPCR) in archived samples from common canaries (*Serinus canaria*) from Germany. Moreover, sequencing of a sample from stump-toed frog (genus *Stumpffia*) collected in northern Madagascar revealed a novel bornavirid that phylogenetically clustered with members of the genus *Cultervirus*.

Materials and methods

Selection of datasets

We compiled a list of datasets using the European Nucleotide Archive Browser's advanced search portal (https://www.ebi.ac.uk/ena/browser/advanced-search), selecting 'raw reads' as the data type. Specifically, this search included all currently recognized taxonomic units across the major vertebrate clades: fish, amphibians, reptiles, and birds as represented in publicly available datasets. We further restricted the search to RNA-derived datasets from metagenomic, transcriptomic, or viral RNA-seq experiments. See Supplementary Materials and Methods for complete search queries.

We also used the Serratus website (Edgar et al. 2022) to identify further datasets within the SRA that potentially contain bornavirus-like sequences. In addition to the publicly available SRA datasets, RNA-seq data from a stump-toed frog was newly sequenced on an Illumina NextSeq instrument using standard laboratory approaches (e.g. Lyra et al. 2021; see Supplementary Materials and Methods). The newly obtained dataset was submitted to SRA (Bioproject PRJNA1418979).

Mining sequence read archive datasets, *de novo* assembly, and identification of novel bornavirid genomes

To detect potential bornavirus-related sequences within the selected SRA datasets, we used 'SRAminer' (https://gitlab.com/FPfaff/sraminer; Pfaff et al. 2023). Following an initial screening of 300 000 read subsets using SRAminer, we selected all datasets with at least one blastx match for more in-depth analysis. We then used *parallel-fastq-dump* (v0.6.7; Valieris 2021) to download the complete datasets for these SRA entries.

The raw reads downloaded from SRA and the stump-toed frog dataset were subsequently trimmed to remove low-quality regions and adapter contamination with *TrimGalore!* (v0.6.10; Krueger F 2019, Martin 2011) in automatic mode. The trimmed reads were subsequently assembled *de novo* using *SPAdes* genome assembler (v4.0.0; Bushmanova et al. 2019), operating in `–rna` mode.

The resulting contigs were subjected to blastx search against a representative public *Bornaviridae* protein database (see Supplementary Data S1) using *diamond blastx* (v2.1.11; Buchfink et al. 2021). Contigs that matched the search criteria were selected for further analysis and imported into Geneious Prime (v2025.1.3) for detailed characterization.

Following the identification of novel bornaviral sequences, the same previous search workflow was repeated using the newly identified viral protein sequences together with public database entries, in order to expand the proteins database and to detect additional related contigs and potentially novel viruses.

To verify the species assignment of the sampled organisms from the SRA datasets, we selected all contigs from the assembly that aligned with the mitochondrial cytochrome B (*MT-CYB*) gene and subjected them to sequence similarity analysis using the NCBI blastn suite (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

Genomic characterization of bornavirid genomes

Potential ORFs were predicted using the 'Find ORFs' tool in Geneious Prime and the deduced amino acid (aa) sequences were compared against the non-redundant NCBI blastp database. Contigs with multiple intact ORFs resembling known bornavirid genes, along

with flanking untranslated regions, were considered as complete bornavirid genomes.

To ensure the accuracy of the genomes, they were screened for any vector or adaptor contamination using the NCBI VecScreen tool (https://www.ncbi.nlm.nih.gov/tools/vecscreen).

Taxonomic classification of novel bornavirid genomes

Given the ambiguous genomic organization of the novel avian carboviruses, we based our initial phylogenetic analysis on an alignment of the N protein. The protein sequences were aligned to reference bornavirid sequences using the Clustal omega algorithm (v1.2.2; Sievers et al. 2011) and refined with Muscle (v3.8.425; Edgar 2004), with both alignments implemented in Geneious Prime. The reference viruses were selected to represent all species within the *Bornaviridae* family recognized by the International Committee on Taxonomy of Viruses (ICTV) (n = 18), and were supplemented with additional bornavirids that are not yet formally classified by ICTV but likely represent variants within existing viral species (n = 5). The phylogenetic relationships were inferred using IQ-TREE (v2.4.0; Minh et al. 2020). The optimal substitution model was selected automatically (−m MFP-MERGE; Kalyaanamoorthy et al. 2017), and branch support was assessed using ultrafast bootstrap (−B; Hoang et al. 2018) and SH-aLRT tests (−alrt), with 1 000 000 replicates for each.

In addition, for better phylogenetic resolution, we also constructed phylogenetic analyses based on the concatenated sequences of the N, G, and L proteins, applying the same IQ-TREE parameters and settings described above. In addition, we applied a partitioning model (−Q; Chernomor et al. 2016) that allowed for distinct substitution models and evolutionary rates within each partition.

Bayesian phylogenetic inference was performed in BEAST for both N protein alone and the concatenated N, G, and L proteins (v1.10.4; Drummond and Rambaut 2007), using an uncorrelated relaxed lognormal clock and constant-size coalescent model. Markov chain Monte Carlo chains were run for 100 000 000 steps, with parameters sampled every 10 000 steps. Effective sample sizes were evaluated using Tracer (v.1.7.2; Suchard et al. 2018), and maximum clade credibility trees were generated with TreeAnnotator (v.1.10.4) after discarding a burn-in of 2 000 000, providing median node heights and posterior support values to complement IQ-TREE analyses.

Finally, a phylogenetic tree based on the G protein sequences was generated using IQ-TREE, following the same workflow described above.

The N and L nt sequences, along with their deduced aa sequences, from all references and previously characterized bornavirids, as well as from the bornaviral sequences identified and selected for further analysis in this study, were initially aligned using the Clustal omega algorithm and subsequently refined with the MUSCLE alignment tool; both alignments were implemented in Geneious Prime. Pairwise distance matrices were subsequently exported and visualized as heatmaps in R (version 4.3.1; R Core Team 2024) using pheatmap package (v1.0.13; Kolde 2010) and RcolorBrewer package (v1.1-3; Neuwirth 2002). The N protein heatmap complements the N phylogenetic tree by visualizing sequence similarity and percent identity among N protein sequences.

Novel viruses were defined based on N gene nt identity: sequences sharing <90% identity with known bornavirids were considered novel, whereas sequences with ≥90% identity retained the corresponding established virus names.

Prediction of transcriptional patterns, sequence motifs and transmembrane domains

The trimmed raw reads corresponding to the newly identified viruses were then mapped back to the assembled genome using the 'bbmap' function from BBTools (version 39.33; Bushnell et al. 2017). Duplicate reads were removed with the 'MarkDuplicates' function from Picard toolkit (version 2.20.4; Broad Insitute 2019), and reads were separated by mapping orientation using Samtools (version 1.22.1; Danecek et al. 2021). The genome coverage for forward and reverse deduplicated reads was calculated using Bedtools (version 2.31.1; Quinlan and Hall 2010) and visualized in R. See Supplementary Materials and Methods for details on transcriptional mapping interpretation and limitations.

Using the trimmed raw reads dataset and corresponding assembled genome, potential introns were predicted with the Geneious Prime generic mapper (v2025.1.3; medium sensitivity; find fusion genes and novel introns). Candidate introns were selected based on the presence of spliced reads supporting the predictions and the presence of canonical `GU-AG` splice site motifs.

Transcriptional initiation and termination sites were predicted based on sequence similarities to known bornaviral sequences and further validated by inspection read coverage patterns, including transitions to poly(A) tail at termination sites. The function 'FIMO' from the MEME Suite (v5.5.2; Bailey and Elkan 1994) was used to identify genus-specific transcriptional regulatory motifs. For the novel culterviruses, the conserved initiation and termination motifs '`GAM`' and '`AKUUAAYAAAAACAU`', respectively, previously described for *Cultervirus* were used (Eshak et al. 2023). The initiation motif '`CCGAA[A/C][A/C]A`' together with the termination motif '`UAA[A/G]AAA`' described for reptilian carboviruses was used to predict transcriptional sites in the novel avian carboviruses identified in this study (Hyndman et al. 2018). The combined use of conserved motifs and coverage-based analyses enabled reliable prediction of transcriptional start and termination sites across the novel bornavirids.

For predicting potential transmembrane domains (TM), signal peptides (SP), and cleavage sites (CS) in the G protein of the new genomes, we used DeepTMHMM (pybiolib; version 1.2.582; Hallgren et al. 2022) and ProP-1.0 (Duckert et al. 2004), respectively.

Discrimination between endogenous bornaviral elements and avian carbovirus genomes

To evaluate whether the identified avian carboviruses represent endogenous bornavirus-like elements (EBLs) integrated into host genomes, we performed nucleotide–nucleotide BLAST searches (BLASTn; version 2.16.0+) using the -task dc-megablast option. Searches were conducted with an E-value threshold of 1E-15 and a minimum nucleotide (nt) identity of 60%. Host genome assemblies corresponding to each representative host species were retrieved from the NCBI Genomes database and used as query sequences. Specifically, genome assemblies bHirRus1.pri.v3, Bayl_Tbic_1.1, ASM1759013v1, Cam_Psub_1.2, and serCan2020 were used for the barn swallow (*Hirundo rustica*), rufous-necked snowfinch (*Pyrgilauda ruficollis*), tawny-flanked prinia (*Prinia subflava*), and common canary (*S. canaria*), respectively. All carboviral sequences identified in this study were used as the BLASTn subject database. BLASTn alignments with aligned regions ≥1000 nucleotides in length and ≥60% nucleotide identity were manually inspected and visualized

using Geneious Prime (v2025.1.3) to assess alignment quality and evaluate potential integration into host genomes.

Specific search for avian carbovirus G and M genes in sequence read archive datasets

For SRA datasets containing avian carboviruses lacking the G gene or both the G and M genes, we specifically examined the data for G- and M-like sequences, given the possibility of segmented genomes. The datasets were downloaded and preprocessed by fastp (v0.23.4) with the options '-l 35 -x -y'. The preprocessed reads were mapped to the corresponding host genomes, and the unmapped reads were extracted. The unmapped reads were assembled by *SPAdes* (v3.15.543) with the default settings and contigs larger than 100 nt were extracted. Sequence similarity searches were performed against a custom database containing protein sequences from the taxonomic group *Orthornavirae* by MMseqs2 (version c48da9d781b81804727b5cccfed7f97cfcc20c9d) using the extracted contigs as queries. The E-value threshold was set to 0.1.

Detection and quantification of co-infecting viruses in avian carbovirus-positive sequence read archive datasets

To identify co-infecting viruses in avian carbovirus-positive SRA datasets, we performed a two-step sequence similarity search. First, we used MMseqs2 (version c48da9d781b81804727b5cccfed7f97cfcc20c9d19) to search contigs from the avian carbovirus-positive datasets (see above) against a custom database of viral protein sequences from the NCBI 'nr' database. Among the sequences that were alignable to each contig, the sequence with the highest score was designated as the top hit, and contigs whose top hits were detected with E-value of less than 10^{-10} were extracted. Next, the extracted contigs were analysed by BLASTx (BLAST+ 2.15.0) against the NCBI 'clustered nr' database (April 2024). The options used were '-evalue 1e-10 -word_size 2'. Each contig was considered evidence for a candidate co-infecting virus if the top hit sequence was a virus, but the following sequences were excluded from the co-infecting virus candidates: (i) if the top hit was a virus belonging to a taxonomic group detected only in non-vertebrate animals, it was excluded from the candidate co-infecting viruses; (ii) if the top hit was a human-derived virus and the sequence identity with the contig was 99% or higher, it was excluded from the co-infected virus candidates because of the possibility of contamination or index hopping during sample preparation and/or sequencing; (iii) kolmioviruses were excluded because they do not encode envelope proteins or capsid proteins; (iv) aligned regions that were low complexity sequences were excluded.

Specific search for canary orthobornaviruses in sequence read archive datasets

Sixty-four RNA-seq datasets derived from common canaries (*S. canaria*; see Supplementary Table S3) were mapped to the genomes of canary bornavirids (CnBV-1 to -3; KC464471.1, KC464478.1, KC595273.2) using Magic-BLAST 1.7.2, and the number of mapped reads for each viral genome was counted using SAMtools (1.19). RNA-seq datasets with 10 or more mapped reads and a read per million (RPM) of 0.1 or higher were considered positive.

Screening of archived samples for avian carboviruses

Archived RNA/DNA originating from tissue or swab samples of 33 common canaries or estrildid finches infected with canary bornavirids (CnBV-1 to -3) or estrildid finch bornavirus 1 (EsBV-1; *Orthobornavirus estrildidae*), respectively, from 19 different flocks were derived from previous studies (Philadelpho et al. 2014, Rubbenstroth et al. 2013, Rubbenstroth et al. 2016). These were tested for canary delgembornavirus (CDBV) RNA using a newly established RT-qPCR assay. Furthermore, RNA was extracted from cell cultures established by inoculation with organ material from three brain samples that were positive for both CDBV and CnBV. As only CnBV-1 or CnBV-3 had previously been confirmed in these cultures, the RNA was additionally tested for the presence of CDBV (Supplementary Table S1). The primers and probes were designed to target the viral N gene of CBDV virus sequence (ERR324847) (see Supplementary Table S2, for complete list of primers and probes). RT-qPCR was performed using the AgPath-ID one-step RT-PCR reagents (Thermo Fisher Scientific). Briefly, 2.5 μl of extracted RNA was reverse transcribed and amplified in a 12.5 μl reaction containing primers CCV-664+ and CCV-815– and probe CCV-715-P at final concentrations of 0.8 and 0.4 μM, respectively. A primer/probe set targeting the cellular ß-actin (*ACTB*) gene was included as an internal control as described previously (Supplementary Table S2) (Wernike et al. 2011). The PCR was run on a Bio-Rad CFX96 qPCR cycler (Bio-Rad) using the following cycling conditions: 45°C for 10 min, 95°C for 10 min, followed by 42 cycles of 95°C for 15 s, 57°C for 20 s, and 72°C for 30 s.

Positive samples were then subjected to conventional RT-PCR using Superscript III One-Step RT-PCR with Platinum *Taq* (Invitrogen). Amplifications were carried out on an Applied Biosystems thermal cycler (Thermo Fisher Scientific) with the following cycling program: 50°C for 30 min, 94°C for 2 min, followed by 40 cycles of 94°C for 30 s, 57°C for 30 s, 68°C for 45 s, with a final extension at 68°C for 5 min. Primers CCV-68+ and CCV-815– (Supplementary Table S2) were used to amplify a product of 747 base pairs (bp) length. RNase-free water was used instead of cDNA as a negative control. PCR amplicons were run on a 2% agarose gel; bands were excised and purified using the QIAquick Gel Extraction Kit (Qiagen). Extracted DNA was submitted for Sanger sequencing using both forward and reverse primers (Supplementary Table S2). Raw sequences were trimmed for quality and sequences of amplification primers were removed before they were assembled and a consensus sequence of 678 bp length was generated for each sample. The obtained sequences were aligned with the *in silico* carboviral sequences and reptilian carboviruses using the Clustal omega algorithm. Alignments were further refined with MUSCLE; both alignments were implemented in Geneious Prime. Phylogenetic relationships were inferred from this final alignment using IQ-TREE as mentioned previously.

To rule out the possibility that the experimentally identified carboviral sequences represented EBLs integrated into the host genome, quantitative PCR (qPCR) was performed on the positive samples. Reactions were prepared using the Maxima Probe/ROX qPCR Master Mix (Thermo Fisher Scientific), according to the manufacturer's instructions. The primers and probes targeting the CDBV N gene and *ACTB* were used as described above (Supplementary Table S2). Amplifications were carried out on a Bio-Rad CFX96 qPCR cycler using the following cycling conditions: 50°C for 2 min, 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, 60°C for 30 s and 72°C for 30 s. As a positive control, a CDBV cDNA template was generated from

positive sample *VS*-4244 using the Superscript III One-Step RT-PCR kit with forward primer CCV-68+ (0.4 μM), treated with RNase Cocktail Enzyme Mix (Invitrogen) at 37°C for 20 min to remove residual RNA, and subsequently purified using AMPure XP beads (Beckman Coulter).

Results

Data mining of sequence read archive datasets

We screened 184 749 publicly available transcriptomic datasets from fish, amphibians, reptiles, and birds (Supplementary Table S3), along with the newly generated RNA-seq dataset from a stump-toed frog, for sequences related to bornavirids. Overall, 501 SRA datasets and the stump-toed frog dataset (SRR37117020), contained at least one read matching a bornavirus reference protein sequence (see Supplementary Table S4, for complete list of the 501 SRA datasets). For each of these datasets, a *de novo* assembly was performed and the resulting contigs were subsequently evaluated. In 268 of the *de novo*-assembled SRA datasets, we identified sequences resembling potential EBLs, which were not analysed further in this study. In 106 SRA datasets, the assembled contigs showed no sequence similarity to the bornavirid reference database, likely resembling false-positive hits. Partial and complete viral genomes from the families *Chuviridae* and *Rhabdoviridae* were detected in three and two datasets, respectively (Supplementary Table S4).

Finally, in 122 *de novo*-assembled SRA datasets, partial or coding-complete bornaviral genome sequences were identified. As some of these SRA datasets represented different organ samples of the same individual, multiple replicates within a single study, or identical cell lines, we selected representative genomes for further characterization. However, for viruses detected in multiple individuals within the same study, more than one representative genome was selected to capture potential inter-host variability (Supplementary Table S4). Four partial and 25 coding-complete bornaviral genomes were selected for further analysis in this study (Table 1). Some of these bornaviral sequences represent already known viruses, while others potentially represent novel viruses that may constitute new viral species and/or genera, as suggested by sequence similarity and phylogenetic analyses.

Overall phylogenetic classification of identified bornaviral genomes

Maximum-likelihood and Bayesian phylogenetic analyses of the predicted viral protein N revealed that the majority of complete and all partial genomes clustered with members of the established genera within the family *Bornaviridae* (Fig. 1). Interestingly, the gold-striped salamander bornavirus (GSSBV) did not cluster together with members of the established genera, branching basal to orthobornaviruses (Fig. 1).

Specifically, 11 complete and 2 partial genome sequences clustered with members of the genus *Orthobornavirus* (Fig. 1 and Table 1). Based on phylogenetic analysis and sequence identity, novel orthobornaviruses were identified from reptiles (n = 6), a bird (n = 1), and a mammal (n = 1). Brown-banded water snake bornavirus (BWBV), marbled whiptail bornavirus (MHBV), striped plateau lizard bornavirus (SPBV), Great Basin rattlesnake bornavirus (GBRBV), prairie rattlesnake bornavirus (PRBV), Southern pacific rattlesnake bornavirus (SPRBV) were detected in reptiles, ochre-bellied flycatcher bornavirus (OFBV) in a bird and Hottentot golden mole bornavirus (HGMBV) in a mammal. Additionally, sequences of the already known viruses parrot bornavirus 2 (PaBV-2; *Orthobornavirus alphapsittaciforme*) and Totonacan rattlesnake bornavirus (TRSBV; *Orthobornavirus caenophidiae)* were detected in a monk parakeet (*Myiopsitta monachus*) and Eastern diamondback rattlesnake (*Crotalus adamanteus*) dataset, respectively (Fig. 1 and Table 1).

Seven complete and two partial genome sequences clustered with members of the genus *Cultervirus* (Fig. 1 and Table 1). Six novel culterviruses were detected in fish (n = 3), reptiles (n = 2), and an amphibian (n = 1). In detail, javeline goby bornavirus (JGBV), Chinese catfish bornavirus (CCBV), and yellow catfish bornavirus (YCBV) were detected in fish, the Asian annulated sea snake bornavirus (ASBV) and tropical rattlesnake bornavirus (TRBV) were both detected in reptiles, and the stump-toed bornavirus (SFBV) was detected in an amphibian dataset. Pará molly bornavirus (PMBV; *Cultervirus poeciliae*) was detected in a green swordtail (*Xiphophorus hellerii*), while sequences of WhSBV were again detected in grass carp-derived cell lines, here identified in hepatocytes, and MDCBV was detected in goldfish (Fig. 1 and Table 1). Partial genome sequences of Bombay duck fish bornavirus (BFBV; *Cultervirus harpadoni*) and the cartilovirus LSBV were too short and excluded from further analysis (Supplementary Table S4).

Six genome sequences clustered closely with members of the genus *Carbovirus* (Fig. 1 and Table 1). From these we propose four novel carboviruses that were detected in avian datasets. These carbovirus genomes were considered complete despite lacking two or three canonical bornaviral ORFs and were designated as barn swallow delgebornavirus (BSDBV), rufous-necked snowfinch delgebornavirus (RSDBV), canary delgembornavirus (CDBV), and tawny-flanked prinia delgembornavirus (TPDBV). The prefixes 'delge' and 'delgem' symbolically refer to the lack of the G ORF ('delta G') and the lack of both the G and M ORFs ('delta G and M'), respectively.

When analysed with BLASTn, the *MT-CYB* gene sequences assembled from each dataset matched the species described in the associated metadata, thereby confirming the identity of the host (data not shown).

Detection of a genetically distinct amphibian bornavirus

The sequence of the newly identified GSSBV, identified from a liver tissue dataset of a gold-striped salamander (*Chioglossa lusitanica*), did not cluster with any of the classified genera of the *Bornaviridae* family (Fig. 1 and Table 1). GSSBV exhibits the canonical orthobornavirus genome organization 3′–N–X/P–M–G–L–5′ and its total genome length is 8845 nt (Fig. 2). Phylogenetic analysis based on the concatenated N, G, and L proteins revealed GSSBV to branch from a node basal to all known orthobornaviruses and may represent a new viral species within a novel genus (Fig. 3). Pairwise comparisons with orthobornaviruses revealed 20%–30% nt and aa identity in the N genes and proteins, and 40%–47% in the L genes and proteins (Supplementary Figs S1 and S2). Similar to *Orthobornavirus* genomes, motif prediction and sequence data mapping revealed three transcriptional start sites (S1–S3) and four termination sites (T1–T4): S1 was located upstream of the N gene; S2 was located upstream of the X/P genes; S3 was located upstream of the G/M/L region (Supplementary Table S5). The termination sites (T1, T2, and T4) were downstream of N, P, and L, respectively, while T3 was found within the L gene. S2 was located upstream of T1, and S3 was positioned adjacent to T2, similar to orthobornaviruses (Supplementary Table S5). The regions between the start and termination sites corresponded to the following predicted transcripts: N

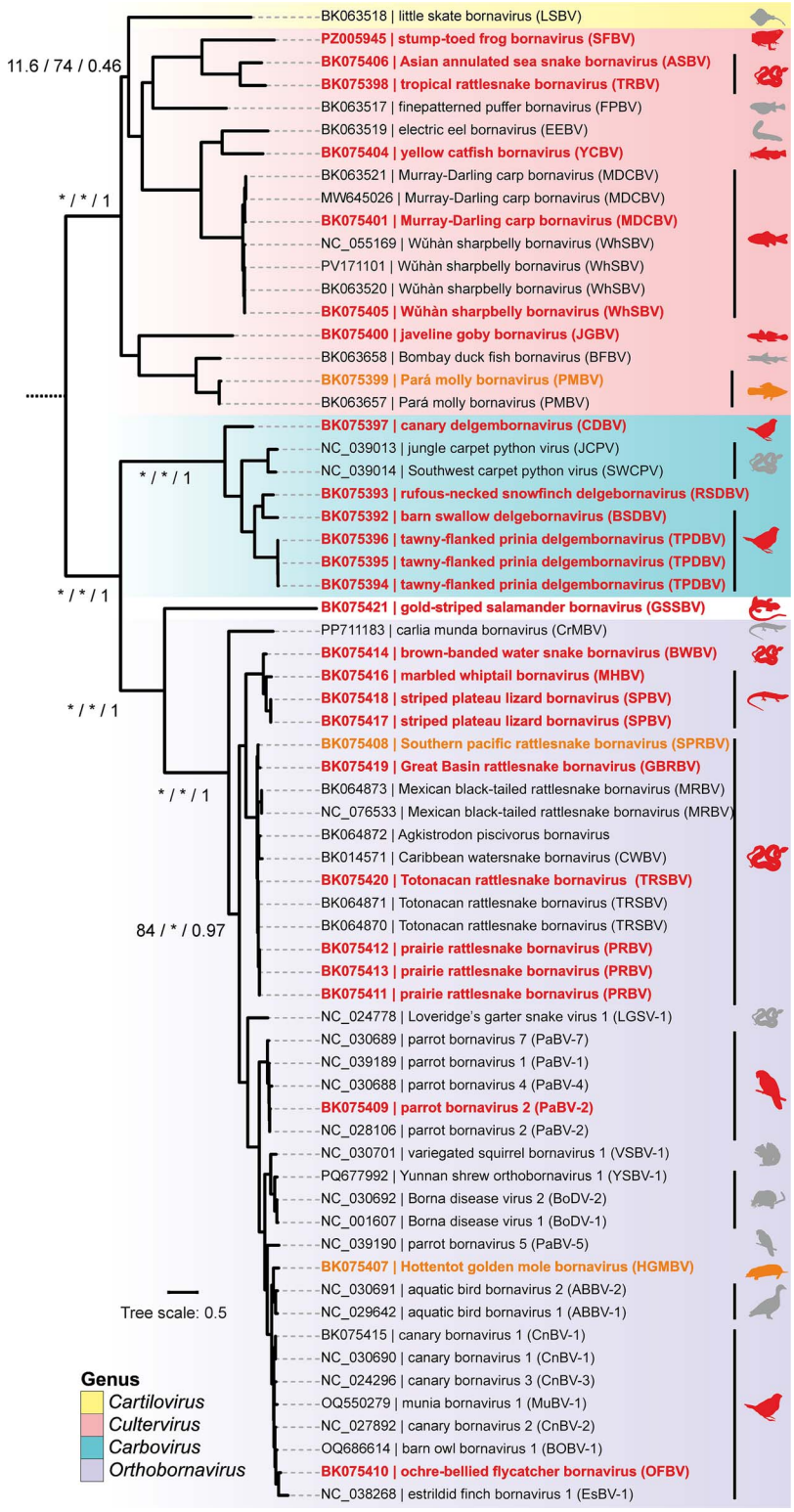

Figure 1 Phylogenetic relationships of the N proteins within the family *Bornaviridae*. The maximum-likelihood tree was based on amino acid sequence alignments of the viral protein N, including newly identified complete bornavirids from fish, amphibians, reptiles, and birds (red text), novel partial bornaviral genomes (orange text), and previously identified members of the genera *Cartilovirus*, *Cultervirus*, *Carbovirus*, and *Orthobornavirus.* Soybean cyst nematode virus 1 (family *Nyamiviridae*; NC_024702.1) was used as an outgroup to root the tree. The silhouettes represent typical host organisms of previously published viruses (highlighted in grey) or the reported sampling source (highlighted in red) of the viral genomes identified in this study. The tree was constructed using IQ-TREE (version 2.4.0) using optimal model section and statistical support with 1 million replicates for ultrafast bootstrap and SH-aLRT test. Branch support for phylogenetic relationships was assessed using a combination of ultrafast bootstrap, SH-aLRT tests in IQ-TREE, and posterior probabilities from an independent Bayesian phylogeny run using BEAST (v1.10.4), providing complementary measures of nodal confidence across maximum likelihood and Bayesian frameworks [ultrafast bootstrap/SH-aLRT/posterior probability]. Asterisks indicate statistical support ≥90% and ≥90% for ultrafast bootstrap and SH-aLRT, respectively. Numbers denote the major branches of the tree.

Table 1 Summary of partial and complete coding bornaviral genomes *de novo* assembled from SRA datasets, which were selected for further characterization.

SRA accession	Sampled organism	NCBI family	Sampled material	Virus detected (abbreviation; GenBank accession)	Genome length (nt)	Virus-mapped reads, n (%)
SRR13516426	Barn swallow *Hirundo rustica* (Linnaeus, 1758)	Hirundinidae	Brain	Barn swallow delge bornavirus (BSDBV; BK075392)	7818	12 574 (0.01)
SRR13995241	Rufous-necked snowfinch *Pyrgilauda ruficollis* (Blanford, 1871)	Passeridae	Blood	Rufous-necked snowfinch delgebornavirus (RSDBV; BK075393)	7902	37 897 (0.002)
SRR16241768	Tawny-flanked prinia *Prinia subflava* (Gmelin,1789)	Cisticolidae	Shell gland	Tawny-flanked prinia delgembornavirus (TPDBV; BK075394)	7943	20 179 (0.03)
SRR16241810	Tawny-flanked prinia *P. subflava* (Gmelin, 1789)	Cisticolidae	Shell gland	Tawny-flanked prinia delgembornavirus (TPDBV; BK075395)	7943	3130 (0.004)
SRR16241812	Tawny-flanked prinia *P. subflava* (Gmelin, 1789)	Cisticolidae	Shell gland	Tawny-flanked prinia delgembornavirus (TPDBV; BK075396)	7943	2951 (0.004)
SRR23747101	Common canary *Serinus canaria* (Linnaeus, 1758)	Fringillidae	Pectoral fatty tissue	Canary delgembornavirus (CDBV; BK075397)	7447	11 688 (0.014)
SRR23747114	Common canary *S. canaria* (Linnaeus, 1758)	Fringillidae	Chest skin	Canary bornavirus 1 (CnBV-1; BK075415)	8934	1 391 610 (1.35)
SRR18056692	Ochre-bellied flycatcher *Mionectes oleaginous* (Lichtenstein, 1823)	Tyrannidae	Pectoralis major skeletal muscle	Ochre-bellied flycatcher bornavirus (OFBV; BK075410)	8936	14 132 (0.03)
SRR13480239	Monk parakeet *Myiopsitta monachus* (Boddaert, 1783)	Psittacidae	Cerebellum	Parrot bornavirus 2 (PaBV-2; BK075409)	8914	208 581 (0.54)
SRR18361051	Prairie rattlesnake *Crotalus viridis viridis* (Rafinesque, 1818)	Viperidae	Kidney	Prairie rattlesnake bornavirus (PRBV; BK075411)	8909	391 296 (0.7)
SRR18361058	Prairie rattlesnake *C. viridis viridis* (Rafinesque, 1818)	Viperidae	Skin	Prairie rattlesnake bornavirus (PRBV; BK075412)	8903	675 933 (1.1)
SRR18361065	Prairie rattlesnake *C. viridis viridis* (Rafinesque, 1818)	Viperidae	Kidney	Prairie rattlesnake bornavirus (PRBV; BK075413)	8910	412 111 (0.7)
SRR27545516	Great Basin rattlesnake *Crotalus lutosus* (Klauber, 1930)	Viperidae	Venom glands and blood	Great Basin rattlesnake bornavirus (GBRBV; BK075419)	8888	154 151 (0.4)
SRR28501301	Eastern diamondback rattlesnake *Crotalus adamanteus* (Palisot De Beauvois, 1799)	Viperidae	Venom gland	Totonacan rattlesnake bornavirus (TRSBV; BK075420)	8886	2364 (0.01)
SRR19582207	Brown-banded water snake *Helicops angulatus* (Linnaeus, 1758)	Dipsadidae	Venom Gland	Brown-banded water snake bornavirus (BWBV; BK075414)	8909	94 853 (0.2)
SRR24893433	Marbled whiptail *Aspidoscelis marmoratus* (Baird & Girard, 1852)	Teiidae	Unknown	Marbled whiptail bornavirus (MHBV; BK075416)	8816	20 715 (0.004)
SRR25583285	Striped plateau lizard *Sceloporus virgatus* (Smith, 1938)	Phrynosomatidae	Abdominal skin	Striped plateau lizard bornavirus (SPBV; BK075417)	8877	8655 (0.02)
SRR27193988	Striped plateau lizard *S. virgatus* (Smith, 1938)	Phrynosomatidae	Liver	Striped plateau lizard bornavirus (SPBV; BK075418)	8877	23 641 (0.2)
SRR12915742	Southern Pacific rattlesnake *Crotalus helleri* (Meek, 1905)	Viperidae	Venom Gland	Southern Pacific rattlesnake bornavirus (SPRBV; BK075408)	1809	N.A.
ERR12667297	Hottentot golden mole *Amblysomus hottentotus* (Smith, 1829)	Chrysochloridae	Heart	Hottentot golden mole bornavirus (HGMBV; BK075407)	1820	N.A.

(*Continued*)

Table 1 Continued

SRA accession	Sampled organism	NCBI family	Sampled material	Virus detected (abbreviation; GenBank accession)	Genome length (nt)	Virus-mapped reads, n (%)
SRR17816101	Gold-striped salamander *Chioglossa lusitanica* (Bocage, 1864)	Salamandridae	Liver	Gold-striped salamander bornavirus (GSSBV; BK075421)	8845	525 101 (0.9)
SRR12915750	Tropical rattlesnake *Crotalus durissus* (Linnaeus, 1758)	Viperidae	Venom Gland	Tropical rattlesnake bornavirus (TRBV; BK075398)	8977	239 680 (0.7)
SRR29672483	Asian annulated sea snake *Hydrophis cyanocinctus* (Daudin, 1803)	Elapidae	Retina	Asian annulated sea snake bornavirus (ASBV; BK075406)	9026	425 870 (0.5)
SRR14361603	Javeline goby *Acanthogobius hasta* (Temminck & Schlegel, 1845)	Gobiidae	Liver	Javeline goby bornavirus (JGBV; BK075400)	8922	6824 (0.14)
SRR25212478	yellow catfish *Tachysurus fulvidraco* (Richardson, 1846)	Bagridae	Liver	Yellow catfish bornavirus (YCBV; BK075404)	9054	7822 (0.02)
SRR25309824	Grass carp *Ctenopharyngodon idella* (Valenciennes, 1844)	Xenocyprididae	Hepatocyte cell line	Wǔhàn sharpbelly bornavirus (WhSBV; BK075405)	9008	96 830 (0.23)
SRR22825171	Goldfish *Carassius auratus* (Linnaeus, 1758)	Cyprinidae	Kidney	Murray-Darling carp bornavirus (MDCBV; BK075401)	8985	30 755 (0.1)
SRR37117020	Sorata stump-toed frog (*Stumpffia sorata*) (Rakotoarison et al. 2017)	Microhylidae	Brain	Stump-toed frog bornavirus (SFBV; PZ005945)	9027	5820 (0.02)
SRR24928263	Chinese catfish *Tachysurus vachellii* (Richardson, 1846)	Bagridae	Spleen, kidney, brain, muscle, ovary, and liver	Chinese catfish bornavirus (CCBV; BK075402–3)	1391; 441	N.A.
SRR12976827	Green swordtail *Xiphophorus helleri* (Heckel, 1848)	Poeciliidae	Fin	Pará molly bornavirus (PMBV; BK075399)	1552	N.A.

N.A.: not applicable.

(monocistronic), X/P (bicistronic) and G/M/L (tricistronic), following the transcription profile pattern of the orthobornaviruses. A single splice site was predicted within the G gene; however, its functional relevance remains unclear (see GSSBV in Fig. 4).

Phylogenetic analysis of the G protein indicates that culterviruses, the cartilovirus LSBV, carboviruses, and GSSBV share a common ancestral lineage. A subsequent lineage split occurred between GSSBV and the carboviruses, while orthobornaviruses form a distinct, independent branch (Fig. 5). Our prediction further indicates that GSSBV encodes a G protein with a putative polybasic CS, an N-terminal signal peptide and transmembrane domains, yielding an overall topology consistent with the canonical structural organization of bornaviral G proteins (see Fig. 5 and Supplementary Table S6).

Characterization of reptilian and avian orthobornaviruses

The novel orthobornaviruses BWBV, MHBV, SPBV, GBRBV, PRBV, TRSBV, OFBV, and MPBV exhibited the characteristic genomic organization 3′-N-X/P-M-G-L-5′ with genome lengths ranging from ∼8800 to 8900 nt (see Fig. 2 and Table 1).

Phylogenetic analyses based on the concatenated N, G, and L protein sequences and pairwise sequence identities at nt and aa levels of the N and L ORFs confirmed the clustering of the newly identified virus genomes together with known orthobornaviruses (see Fig. 3 and Supplementary Figs S1–S2). Among them, BWBV, MHBV, and SPBV formed a distinct phylogenetic clade, sharing 74%–75% nt identity in the N gene among each other, but only 51%–62% nt identity with other known orthobornaviruses. In contrast, GBRBV, PRBV, and TRSBV clustered closely with reptilian orthobornaviruses, showing the highest N and L gene and protein identities >75% to members of the species *Orthobornavirus caenophidiae*. The novel avian orthobornavirus OFBV clustered with members of the species *Orthobornavirus serini*, showing 77%–80% N nt identity with them.

Partial sequences of the novel orthobornavirus SPRBV, comprising a complete X gene and partial N and P genes, clustered with members of *Orthobornavirus caenophidiae* and shared 79%–85% N gene identity (see Fig. 1 and Supplementary Fig. S1). For HMGBV, complete N and X genes and a partial P gene were detected. The N sequence exhibited 54%–76% nt identity with orthobornaviruses, with aquatic bird bornavirus 1 and 2 (ABBV-1 and -2; *Orthobornavirus avisaquaticae*) being the most closely related viruses, showing 75 and 74% N nt identity, respectively (see Fig. 1 and Supplementary Fig. S1).

The novel orthobornaviruses broadly follow the transcriptional strategy previously described for BoDV-1 and other known orthobornaviruses, with three putative transcription initiation sites (S1–S3) and four termination sites (T1–T4). In detail, transcription start sites S1, S2, and S3 are located upstream of N, X/P, and G/M/L, respectively.

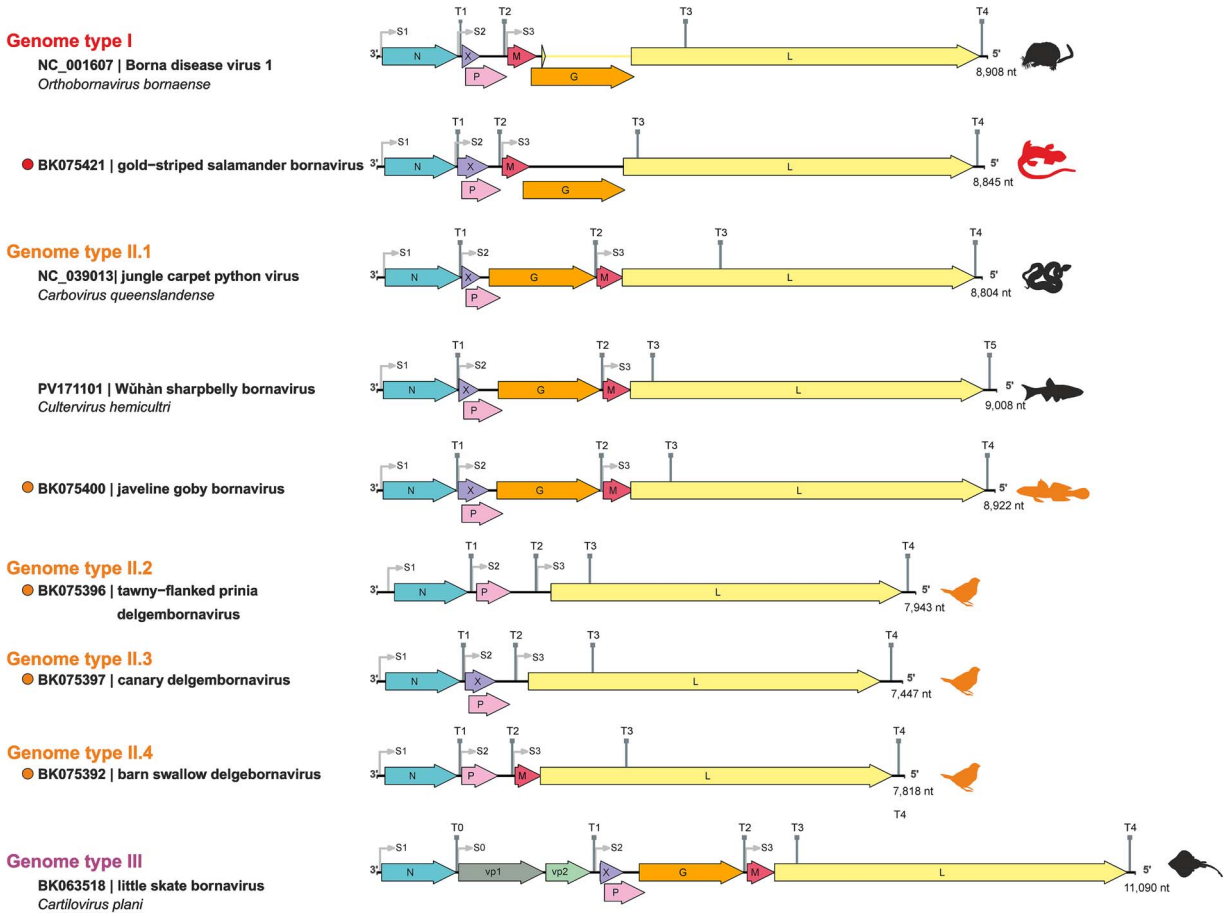

Figure 2 Genome organization patterns of previously described and novel bornaviruses. Representative genome organizations are shown for selected representative viruses alongside putative novel viruses (indicated by coloured dots). Predicted ORFs are shown as arrows along with their associated predicted transcription start (S) and termination (T) sites. The potential host/source of each virus genome is also indicated. Complete *Bornaviridae* genomes typically encode six canonical complete ORFs arranged as follows: 3′-N-X/P-M-G-L-5′ (genera *Orthobornavirus* and GSSBV), 3′-N-X/P-G-M-L-5′ (genera *Carbovirus* and *Cultervirus*), and 3′-N-vp1-vp2-X/P-G-M-L-5′ (genus *Cartilovirus*). Among the novel avian carboviruses, three distinct organization patterns of genome type II were observed, each lacking two to three of the six canonical ORFs: 3′-N-P-L-5′, 3′-N-X/P-L-5, and 3′-N-P-M-L-5′.

Termination sites T1, T2, and T4 are positioned downstream of N, P, and L, respectively, and T3 is located inside the L ORF (Supplementary Table S5). BoDV-1 contains three characteristic introns in the G/M/L region. Among the newly identified viruses, this full intron architecture was retained by OFBV (Fig. 4), whereas the squamate orthobornaviruses showed only two introns.

Our prediction further indicates that all novel orthobornaviruses encode a G protein with a putative polybasic CS, an N-terminal signal peptide and transmembrane domains, yielding an overall topology consistent with the canonical structural organization of bornaviral G proteins (Fig. 5 and Supplementary Table S6).

Characterization of fish, reptilian, and amphibian culterviruses

The assembled genome sequences of the novel culterviruses JGBV, YCBV, ASBV, TRBV, and SFBV are between 8900 and 9000 nt in length and follow the canonical *Cultervirus* genome organization 3′-N-X/P-G-M-L-5′ (Fig. 2 and Table 1).

Phylogenetic analyses based on the concatenated N, G, and L protein sequences demonstrated the reptilian and amphibian viruses ASBV, TRBV, and SFBV to form a monophyletic group (Fig. 3). Pairwise sequence identities of N and L nt and aa sequences, indicated substantial sequence divergence among culterviruses. The newly detected fish cultervirus JGBV shared 33%–43% N gene identity with other culterviruses, while YCBV shared 35%–53%. The reptilian ASBV and TRBV showed 62% N nt identity with each other but only 36%–48% identity to culterviruses. The amphibian virus SFBV shared 35%–48% N nt identity with culterviruses (Supplementary Figs S1 and S2).

For the partial genome of the novel cultervirus CCBV, only a partial G ORF and a complete M ORF were available. Phylogenetic analysis showed that the G protein sequence clustered with the newly identified YCBV (Fig. 5), sharing ∼82% and 83% nt identity for G and M, respectively. In the partial genome of PMBV from green swordtail, N and partial X and P ORFs were detected, with the N gene sharing 97% nt identity to the PMBV (BK063657; Fig. 1 and Supplementary Fig. S1).

The novel cultervirus genomes generally exhibit the transcriptional strategy and motifs previously reported for culterviruses, with three major start sites (S1–S3) and four termination sites (T1–T4; Supplementary Table S5). S1, S2, and S3 are located upstream of N, X/P/G, M/L, respectively. Termination sites T1, T2, and T4 are positioned downstream of N, G, and L, respectively, and T3 is located inside the L ORF. S2 and S3 are located immediately adjacent to T1 and T2, respectively (see JGBV and TRBV in Fig. 4).

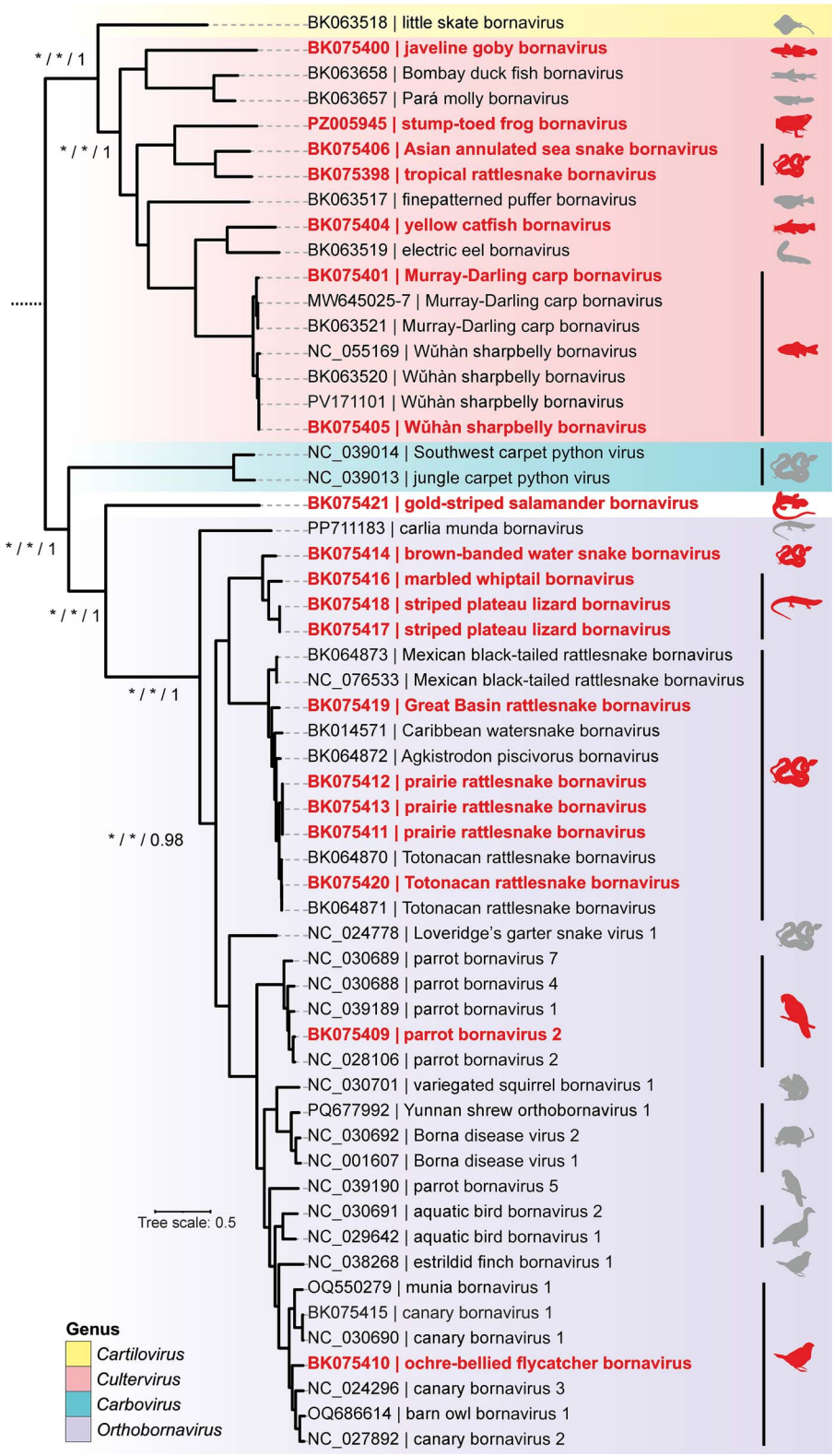

Figure 3 Phylogenetic relationships within the family *Bornaviridae*. The maximum-likelihood tree was based on the concatenated amino acid sequence alignments of the viral proteins N, G, and L of the newly identified bornavirids (all shown in bold red) together with the previously identified members of the genera *Cartilovirus*, *Cultervirus*, *Carbovirus*, and *Orthobornavirus* (bold black). Two viruses of the family *Nyamiviridae* [Soybean cyst nematode virus 1 (NC_024702.1) and Nyamanini virus (NC_012703.1)] were used as an outgroup to root the tree (data not shown). The silhouettes represent typical host organisms of previously published bornavirids (highlighted in grey) or the reported sampling source (highlighted in red) of the viral genomes identified in this study. The tree was constructed using IQ-TREE (version 2.4.0), using optimal partition model selection and statistical support with 1 million replicates for ultrafast bootstrap and SH-aLRT test. Branch support for phylogenetic relationships was assessed using a combination of ultrafast bootstrap, SH-aLRT tests in IQ-TREE, and posterior probabilities from BEAST, providing complementary measures of nodal confidence across maximum likelihood and Bayesian frameworks [ultrafast bootstrap/SH-aLRT/posterior probability]. Asterisks indicate statistical support ≥90% and ≥90% for ultrafast bootstrap and SH-aLRT, respectively. Numbers are shown only for the major branches of the tree.

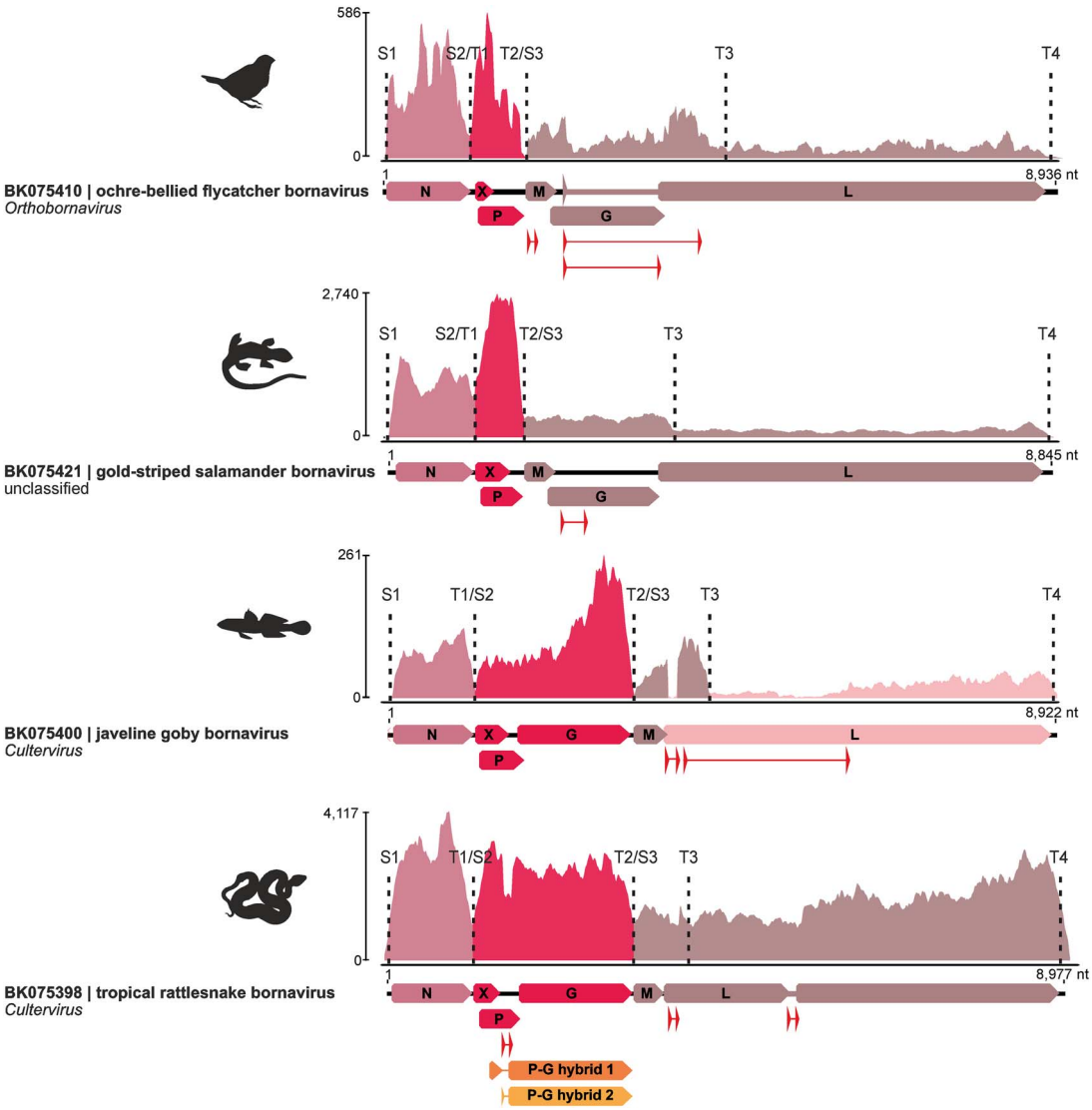

Figure 4 Transcriptional profiles of novel bornavirids. Trimmed raw metagenomic RNA reads were mapped to bornaviral genomes using 'BBmap' (version 39.33). Schematic genome organizations are shown with ORFs represented as arrows, transcription start (S1–S3) and termination (T1–T4) sites are indicated by black dashed lines. ORFs predicted to be co-transcribed on the same RNA transcript are shown in the same colour. Potential introns and intron-derived products predicted using Geneious Prime generic mapper (v2025.1.3; medium sensitivity; find fusion genes and novel introns) are represented by red arrows and orange ORFs, respectively.

A conserved splice site at the start of the L ORF was observed in all the novel culterviruses, potentially enabling expression of M independently from L by utilizing T3. Retention of this intron allows a longer M–L transcript to be expressed, though with lower coverage than the M transcript (see JGBV in Fig. 4). For the reptilian cultervirus TRBV, no significant coverage difference was observed between M and L, despite the presence of the splice site in its genome (see TRBV in Fig. 4). In some newly identified viruses, we detected multiple putative introns of varying lengths within the L ORF, including in JGBV, TRBV, and ASBV (see JGBV and TRBV in Fig. 4).

Our prediction reveals that all novel fish culterviruses encode a G protein with a putative polybasic CS, an N-terminal signal peptide and transmembrane domains, yielding an overall topology consistent with the canonical structural organization of bornaviral G proteins (see WhSBV in Fig. 5 and Supplementary Table S6). Reptilian culterviruses (TRBV and ASBV) also possess a polybasic CS but deviate from the canonical G protein organization, as the N-terminal signal peptide is absent (see TRBV in Fig. 5 and Supplementary Table S6). The amphibian SFBV G protein, however, lacks a CS but retains the canonical structural features characteristic of bornaviral G proteins (see Supplementary Table S6).

Interestingly, in the genomes of the novel reptilian and amphibian culterviruses ASBV, SFBV, and TRBV we detected an intron within the P ORF, i.e. absent in all known fish culterviruses (see TRBV in Fig. 4). This intron potentially gives rise to one or two hybrids of G and P ORFs, provisionally designated as P-G hybrid 1 and 2 (see TRBV in Fig. 4). P-G hybrid 1 is the larger ORF, initiating from a start codon within the X ORF and terminating at the G ORF stop codon. It encodes a longer protein beginning with an intracellular helix followed by a transmembrane helix. P-G hybrid 2 is comparably smaller, starting in the last quarter of the P gene and ending also at the same stop codon of the G ORF (see TRBV in Figs 4–5). P-G hybrid 2 retains the canonical G

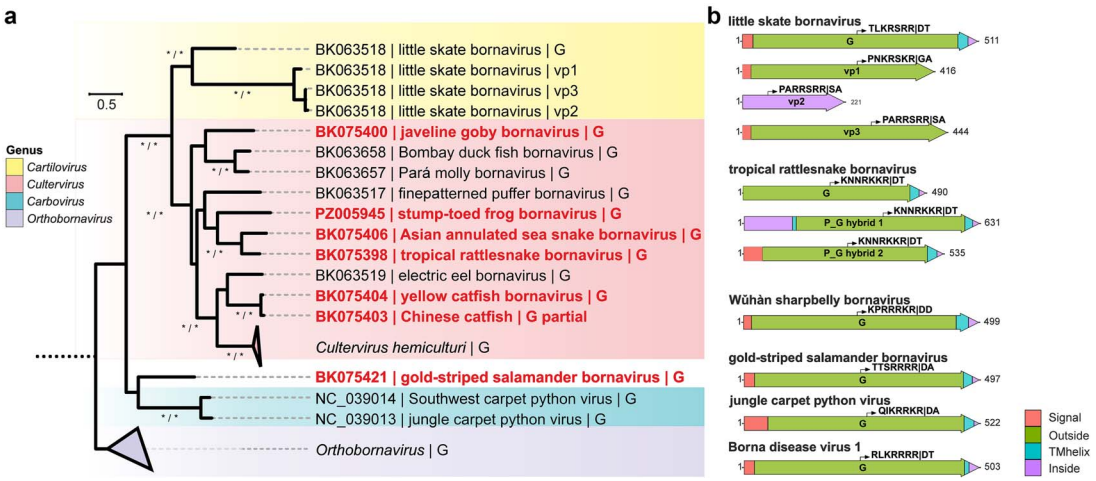

Figure 5 Phylogenetic analysis and structural predictions of bornavirid glycoproteins. (a) Maximum-likelihood phylogeny was based on amino acid sequence alignments of glycoproteins (G) from newly identified bornavirids in fish, amphibians, reptiles, and avian hosts (red text). Previously identified members of the genera *Cartilovirus*, *Cultervirus*, *Carbovirus*, and *Orthobornavirus* were included. A member of the family *Nyamiviridae* (Soybean cyst nematode virus 1; NC_024702.1) were used as outgroup taxa (data not shown). The tree was constructed using IQ-TREE (version 2.4.0) and statistical support with 1 million replicates for ultrafast bootstrap and SH-aLRT test. Branch support for phylogenetic relationships was assessed using a combination of ultrafast bootstrap, SH-aLRT tests [ultrafast bootstrap/SH-aLRT]. Asterisks indicate statistical support ≥90% and ≥90% for ultrafast bootstrap and SH-aLRT, respectively. (b) Predicted glycoprotein architecture of canonical G and putative spliced G variants. Potential signal peptides (SP), transmembrane domains (TM), and furin cleavage sites (CS) (all shown as coloured bars) were predicted using DeepTMHMM (pybiolib; version 1.2.582) and ProP-1.0, respectively.

architecture of bornavirids, including the N-terminal signal peptide that was absent from the G protein and the polybasic CS. The P-G hybrids 1 and 2 of the amphibian SFBV retained the structural features of the reptilian culterviruses, including a bibasic CS that were absent from the canonical short G protein (Supplementary Table S6).

Characterization of avian carboviruses with unique genome organizations

The novel avian carboviruses BSDBV, RSDBV, TPDBV, and CDBV clustered with the two reptilian members of the genus *Carbovirus* based on the N protein phylogenetic analysis (Fig. 1), confirming their placement within the same genus. Pairwise distance analysis of the N and L indicated overall nt and aa identities of 50%–60% between avian and reptilian carboviruses. Among the avian carboviruses, BSDBV and RSDBV were most closely related, sharing 69% nt and 74% aa identity in the N gene and protein (see Supplementary Figs S1 and S2).

Surprisingly, the genome structure of these novel avian carboviruses did not conform to the canonical *Carbovirus* genome organization of 3′-N-X/P-G-M-L-5′. Instead, they lacked two or three of the six canonical bornaviral ORFs, with the G gene absent in all of them and the M and X genes additionally missing in some. Consequently, three organization patterns were observed among these sequences: (i) 3′-N-P-L-5′ for TPDBV (7943 nt), (ii) 3′-N-X/P-L-5′ for CDBV (7447 nt), and (iii) 3′-N-P-M-L-5′ for BSDBV and RSDBV (7800–7900 nt; Fig. 2 and Table 1).

BLASTn searches were conducted against the genomes of the avian hosts from which the corresponding SRA entries had been derived, in order to determine whether these sequences represented recently integrated EBLs. No corresponding hits were identified in the available host genomes, suggesting that these carboviral sequences likely represent exogenous viruses rather than integrated EBLs.

The novel avian carboviruses exhibited three putative transcriptional start sites (S1–S3) and four termination sites (T1–T4; see Supplementary Table S5). Transcriptional motifs known from reptilian carboviruses were readily identified in BSDBV, but several of the described motifs could not be detected in RSDBV, TPDBV, and CDBV based on sequence motif searches. Nevertheless, the transcriptional coverage profile indicated that S1, S2, and S3 are located upstream of N, P or X/P, and L or M/L, respectively. Termination sites are positioned downstream of N (T1), P (T2), and L (T4), as well as within L (T3). Notably, S2 and S3 are located immediately adjacent to T1 and T2, respectively (Fig. 6).

Avian carboviruses sharing the same genomic architecture also exhibited similar transcriptional profiles. In TPDBV and BSDBV, N and P are each expressed from monocistronic mRNAs, whereas in CDBV, X and P are co-expressed from bicistronic mRNA. In TPDBV and CDBV, the L ORF is likely expressed from a monocistronic mRNA, while in BSDBV, M and L appeared to be expressed from a bicistronic mRNA (see Fig. 6).

Predicted splicing patterns revealed a conserved splice site. In BSDBV and RSBDV, the intron is located between the P and M ORFs, whereas in TPDBV the intron is positioned between the P and L ORFs. This intron was not detected in CDBV, which is the only avian carbovirus that retains both the X and P ORFs. Further analysis showed that this splice site generates a longer P isoform, with a size 615 nt in BSDBV and RSBDV and 582 nt in TPDBV (Fig. 6).

Co-detection of viruses in avian carbovirus-positive datasets

The newly identified avian carboviruses lack canonical G and/or M genes, raising the fundamental question of how these viruses achieve host cell entry, assembly, and mediate transmission in the apparent absence of structural proteins typically required for these processes. Although no segmented bornavirus has so far been reported, the possibility that some carboviruses may be segmented cannot be excluded. To assess whether the missing G and M genes might

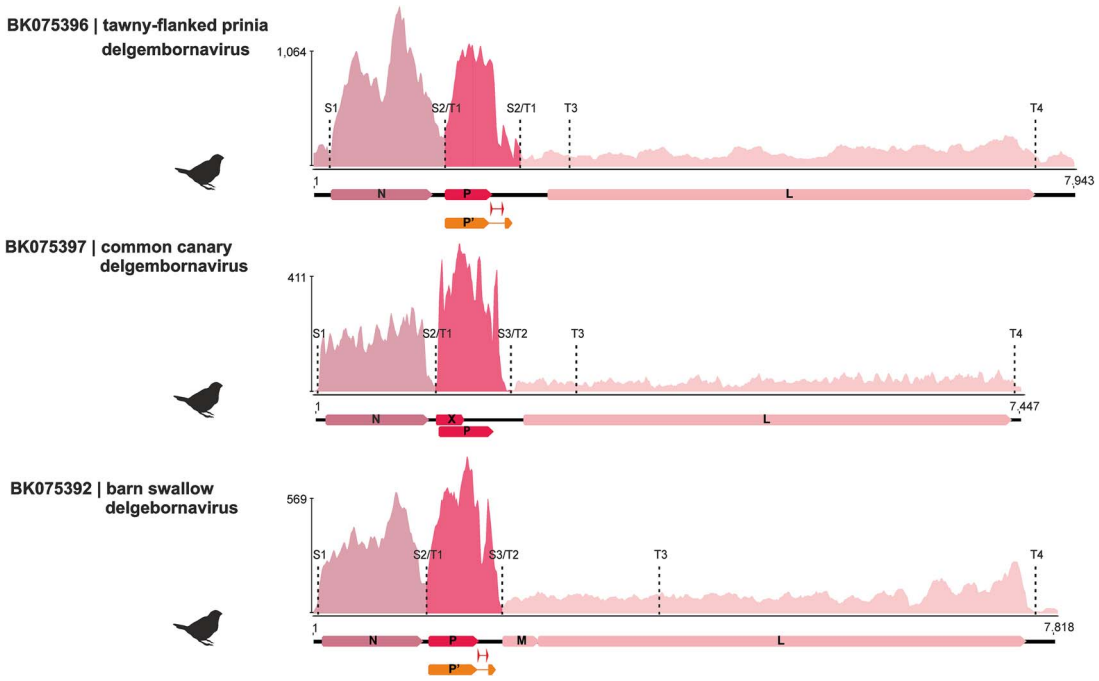

Figure 6 Transcriptional profiles of the novel avian carboviruses in avian hosts. Trimmed raw metagenomic RNA reads were mapped to avian carbovirus genomes using 'BBmap' (version 39.33), with forward- and reverse-oriented reads combined for visualization. Schematic genome organizations are displayed with predicted ORFs represented as coloured arrows. Transcription start (S1–S3) and termination (T1–T4) sites are indicated by black dashed lines. ORFs predicted to be co-transcribed on the same RNA transcript are shown in the same colour. Potential introns and introns-derived products predicted using Geneious Prime generic mapper (v2025.1.3; medium sensitivity; find fusion genes and novel introns) are represented by red arrows and orange ORFs, respectively.

be present on separate genome segments, avian carbovirus-positive datasets were re-analysed using *de novo* assembly and sequence similarity searches against a viral protein database. However, no contigs showing similarity to known bornaviral G or M proteins were identified.

As no contigs related to known bornaviral G or M proteins were detected, avian carbovirus-positive datasets were subsequently screened for additional co-infecting viruses. Ten out of seventeen avian carbovirus-positive datasets contained contigs with similarity to viruses from the families *Bornaviridae*, *Polyomaviridae*, *Astroviridae*, *Parvoviridae*, *Adenoviridae*, *Hepadnaviridae*, *Flaviviridae*, *Picornaviridae*, and *Peribunyaviridae* (Supplementary Table S7). Interestingly, in 2 of 3 CDBV-positive datasets, CnBVs were co-detected (Supplementary Table S7).

Because CnBVs were co-detected with CDBV in common canary datasets, 64 selected common canary-derived datasets were further examined by targeted read mapping to improve the sensitivity of CnBV detection. This approach detected CnBVs in 12 out of 14 (85.7%) CDBV-positive datasets and in 36 of 50 (72.0%) CDBV-negative datasets (Supplementary Fig. S3 and Supplementary Table S8).

Detection of canary delgembornavirus in avian samples

Since CDBV was detected together with CnBVs in datasets from common canaries, archived tissue or swab samples derived from common canaries or estrildid finches in Germany in 2011 to 2015, which had been previously confirmed to be positive for CnBVs or EsBV-1, respectively (Philadelpho et al. 2014, Rubbenstroth et al. 2013, Rubbenstroth et al. 2016), were screened using an RT-qPCR assay targeting the CDBV N gene. In total, 15 out of 30 (50%) CnBV-positive canary samples from 11 different flocks tested also positive for CDBV with Cq values ranging from 15.6 to 29.2 (Supplementary Table S1). The positive samples included brain, pooled organ samples and combined oropharyngeal/cloacal swabs. In contrast, none of the three brain samples from EsBV-1-infected estrildid finches tested positive (Supplementary Table S1). CDBV RNA was not detected in three CnBV cell culture isolates that had been derived from CDBV and CnBV positive brain samples.

To test whether the carboviral sequences from these samples might represent EBLs, qPCR without prior RT step was performed on 12 selected CDBV-positive samples. While the avian carboviral cDNA control yielded a Cq of 24, none of the samples' RNA/DNAs were positive (data not shown).

A phylogenetic analysis of partial N gene sequences (678 nt) showed that all the *in silico*-derived CDBV sequences and those obtained from archived samples clustered together, sharing 96%–100% nt identity (see Fig. 7).

Discussion

In silico screening has become an increasingly important approach for identifying previously unrecognized diversity within the *Bornaviridae* family (Eshak et al. 2023, Ji et al. 2024, Pfaff and Rubbenstroth 2021). Using this approach, this study expands the known spectrum of bornavirids by identifying novel viruses that exhibit previously unrecognized host associations, transcriptional features and genomic

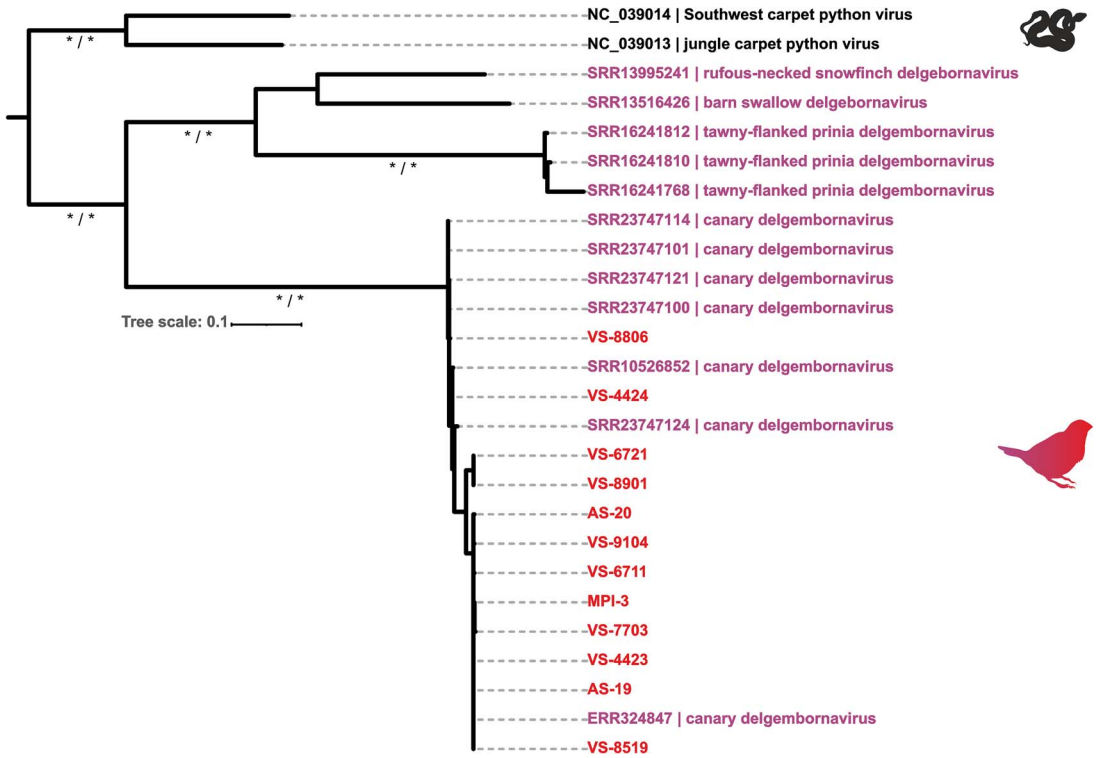

Figure 7 Phylogenetic relationships of carboviruses from avian and reptilian hosts. Maximum-likelihood phylogeny was based on nucleotide sequence alignments of a 678 nt alignment of the nucleoprotein (N) coding gene. Carbovirus sequences identified *in silico* (purple) and detected from available avian samples (red) are shown alongside reference reptilian carboviruses, including jungle carpet python virus (NC_039013) and Southwest python virus (NC_039014), shown in black. The tree was constructed using IQ-TREE (version 2.4.0) and rooted with the reference reptilian carboviruses, an optimal partition model and statistical support with 1 million replicated each for ultrafast bootstrap and SH-aLRT test. Branch support for phylogenetic relationships was assessed using a combination of ultrafast bootstrap and SH-aLRT tests [ultrafast bootstrap/SH-aLRT]. Asterisks indicate statistical support ≥90% and ≥90% for ultrafast bootstrap and SH-aLRT, respectively.

architectures. Beyond the detection of exogenous viruses, genomic screening can also provide access to EBLs, which represent molecular records of ancient bornaviral infections. Although EBLs were excluded from the phylogenetic analyses presented here because most are highly fragmented and degraded, relatively recent or intact EBLs may nevertheless provide valuable insights into bornaviral host range evolution and ancestral genomic content. They therefore represent an important focus for future studies.

The novel viruses identified here, however, provide direct evidence for extant bornaviral diversity and enable more robust phylogenetic placement. In this context, three newly identified reptile-associated viruses grouped within the genus *Orthobornavirus*. Their relatively low N and L gene identity to known reptilian orthobornaviruses supports their classification as putative new species, although formal taxonomic assignment will require ICTV evaluation. All newly identified orthobornaviruses exhibited transcriptional features consistent with those observed in BoDV-1 (Tomonaga et al. 2002, Tomonaga et al. 2000). All three introns previously described in the BoDV-1 M, G and L ORFs (Tomonaga et al. 2002, Tomonaga et al. 2000), were detected in the avian orthobornaviruses (Fig. 4). In contrast, the newly identified squamate orthobornaviruses utilized only two introns. This discrepancy may be due to limitations in the *in silico* prediction or incomplete sequence coverage. However, it could also suggest genuine lineage-specific differences in transcriptional regulation among orthobornaviruses.

Identification of GSSBV expands the known host range of the *Bornaviridae* family to include amphibians. Although the genome organization and inferred transcriptional features of GSSBV are similar to those of orthobornaviruses, phylogenetic analyses and sequence identity comparisons indicate that GSSBV is clustering outside or basal to the genus. These observations indicate that GSSBV may represent a divergent lineage within the family *Bornaviridae*, pending further characterization.

Interestingly, GSSBV exhibits a genome organization that closely resembles that of orthobornaviruses, yet phylogenetic analysis based solely on the G protein places it basal to carboviruses (Fig. 5). Given that carboviruses share a genomic architecture with culterviruses, this incongruence between gene-specific phylogeny and overall genome organization suggests that reordering of the G and M ORFs may have occurred independently on multiple occasions during bornaviral evolution, rather than resulting from a single ancestral rearrangement event. While the directionality of these events remains uncertain, such lineage-specific reorganizations could reflect differing evolutionary constraints. Collectively, these observations highlight the distinct evolutionary position of GSSBV relative to established orthobornaviruses.

Furthermore, our analyses broaden the host range of the genus *Cultervirus* to now also include amphibians and reptiles (Rubbenstroth et al. 2021). Although the detection of culterviruses in amphibians and reptiles may seem unexpected, it may reflect shared evolutionary ancestry and partially overlapping host ecology within this

virus group. The observation that WhSBV can replicate to a limited extent in reptile-derived cells supports the possibility that culterviruses retain conserved biological properties across divergent vertebrate hosts (Eshak et al. 2025). Thus, these findings may indicate a broader host-associated evolutionary history rather than isolated spillover events. However, further sampling and experimental studies are needed to determine whether amphibians and reptiles are true hosts or represent incidental detections.

Transcription of the newly identified culterviruses follows the previously described conserved profiles and motifs (Eshak et al. 2023). A defining feature shared by both newly and previously identified culterviruses is an intron at the beginning of the L gene, which may play a role in regulating the balance between the short M transcript and the longer M–L transcript. In WhSBV, the presence of this intron and several additional introns in the L ORF were experimentally confirmed (Eshak et al. 2025, Eshak et al. 2023). In WhSBV, retention of these introns disrupts the L reading frame and would therefore prevent expression of an intact L protein. In the newly identified culterviruses, however, the situation appears different: removal of the identified introns by splicing seems to be required to maintain the L ORF in frame. In this respect, these culterviruses resemble orthobornaviruses, in which splicing is likewise necessary to generate a functional L transcript. However, differences in intron position and transcriptional context suggest that culterviruses may have implemented this mechanism in a distinct manner, potentially contributing to lineage-specific regulation of L expression or translation. This study further indicates the presence of lineage-specific splicing strategies. Conserved splice sites were identified within the P gene of reptilian culterviruses, including TRBV and ASBV, and the amphibian cultervirus SFBV. These splice sites may facilitate expression of the complete canonical G protein and P–G hybrid proteins of currently unknown function (Fig. 5).

The most striking finding of this study was the identification of new viruses within the phylogenetic cluster of carboviruses, which possess a unique genome organization (Figs 1–2). Notably, these putative carboviruses lack the canonical G ORF and in some cases also X and/or M ORFs.

In bornavirids, the M protein is known to play a critical role in virion assembly and budding by bridging the ribonucleoprotein (RNP) complex with the viral envelope and G proteins (Neumann et al. 2009). The importance of M has been demonstrated experimentally for BoDV-1, where recombinant viruses lacking M (ΔM) or both M and G (ΔMG) are unable to produce infectious progeny particles and become non-transmissive unless the missing proteins are supplied (Fujino et al. 2017). Likewise, the glycoprotein of bornavirids plays a crucial role as a class-III fusion protein that mediates receptor binding and pH-dependent membrane fusion, functions essential for viral entry and spread (Garry and Garry 2009; Rubbenstroth 2022). Exogenous expression of G, together with M, was shown to markedly enhance infectious particle production during BoDV-1 rescue, underscoring that both proteins are indispensable for productive infection (Kanda et al. 2022). The X protein of BoDV-1 plays essential roles in virus replication and persistence. The X protein interacts with P (Wolff et al. 2000), a critical cofactor of the viral RdRp complex (Poenisch et al. 2009). Through this interaction, X functions as key regulatory of viral RNA synthesis and the proper assembly of the polymerase complex (Poenisch et al. 2008, Poenisch et al. 2004). This role appears to be essential, since recombinant BoDV-1 variants carrying either an inactivated X gene or mutations disrupting the X–P interaction cannot be rescued, demonstrating that X is indispensable for viral replication (Poenisch et al. 2007). Beyond its involvement in the replication machinery, X also fulfils an important host-modulatory function. It localizes to mitochondria, where it effectively blocks apoptosis in infected cells, thereby supporting the establishment and maintenance of persistent infection (Poenisch et al. 2009).

Given these important functions in BoDV-1, the absence of X, M, and G ORFs in the avian carboviruses suggests that these novel carboviruses are unlikely to complete assembly, budding, or cell-to-cell spread independently. This raises the hypothesis that they may rely on helper functions provided by co-infecting enveloped viruses, to support replication or particle formation.

Virus–virus interactions and co-infections are increasingly recognized as significant factors in viral evolution, genome complementation, and host adaption (Acchioni et al. 2024, González Aparicio et al. 2022). A well-established example of viral dependence on a helper virus is the hepatitis D virus (HDV; *Deltavirus italiense*; family *Kolmioviridae*), a negative-sense single-stranded RNA virus. HDV is considered a defective virus, as it encodes only one protein, the hepatitis delta antigen, and lacks genes for essential viral functions, including the RdRp required for replication and the envelope G proteins necessary for virion assembly and egress. Consequently, HDV is entirely dependent on co-infection with hepatitis B virus (HBV; *Orthohepadnavirus hominoidei*; *Hepadnaviridae*), which supplies the hepatitis B surface antigen (HBsAg) used to form infectious HDV particles (Lempp and Urban 2017, Sureau and Negro 2016, Taylor 2015). However, deltaviruses are not strictly dependent on HBV alone. Experimental evidence demonstrates that unrelated enveloped viruses can act as helper viruses for HDV, with heterologous envelope G proteins enabling RNP packaging and dissemination (Perez-Vargas et al. 2019). In addition, snake deltavirus has been shown to utilize the envelope proteins of reptarena- and hartmaniviruses (*Arenaviridae*) to generate infectious particles, further illustrating helper-virus flexibility (Hetzel et al. 2019, Szirovicza et al. 2020). Together, these observations indicate that gene-deficient RNA viruses may exploit a range of co-infecting enveloped viruses for assembly and spread.

By analogy, avian carboviruses that lack canonical genes such as G, X, and M may depend on co-infection with other viruses to provide the missing structural or functional components needed for successful replication or transmission. Consistent with this hypothesis, each of the novel avian carboviruses identified in this study was co-detected with at least one additional virus in at least one dataset. However, co-detection was not observed consistently across all datasets, as several avian carbovirus-positive datasets yielded no additional viral hit. Notably, CDBV was the only carbovirus co-detected with enveloped viruses from the same family, specifically CnBVs, suggesting that potential enveloped helper viruses, where present, may not necessarily be restricted to members of the *Bornaviridae*.

The detection of CDBV in a subset of previously CnBV-positive canary samples, together with the frequent co-detection of CnBVs in CDBV-positive RNA-seq datasets (12/14; 85.7%), suggests that CDBV may commonly occur in the context of orthobornavirus co-infections in canaries. However, this association should be interpreted with caution. CnBVs were not detected in all CDBV-positive datasets, and the experimental screening was limited to previously bornavirus-positive samples, precluding conclusions about the occurrence of CDBV in CnBV-negative canaries. Thus, while these observations are compatible with a co-infection scenario, they do not demonstrate an exclusive or obligate association between CDBV and CnBVs.

These observations underscore the need for additional studies to determine whether these avian carboviruses require a helper virus for replication and transmission. Interestingly, the RSDBV, BSDBV, and TPDBV contain introns within the P ORF that give rise to a longer P isoform, which may potentially compensate for the absence of the X ORF. However, experimental validation is required to determine the function relevance and biological of this splice site. Importantly, our findings strongly suggest that the *in silico*- and experimentally identified carboviral sequences are unlikely to represent EBLs within the host genome, and are instead more likely to represent exogenous viruses. However, their pathogenic potential and evolutionary role within the *Bornaviridae* family remain to be determined.

Conclusion

Continuous *in silico* screening of publicly available datasets continues to uncover hidden diversity within the family *Bornaviridae*. Here, we identified previously unrecognized orthobornaviruses, culterviruses, and carboviruses in novel host contexts, including amphibians and reptiles, thereby expanding the known host range and evolutionary diversity of bornavirids. The newly identified viruses contain both conserved features consistent with previously characterized bornavirids, such as BoDV-1 and WhSBV, and lineage-specific genomic traits, including distinct transcriptional motifs, splice sites, and glycoprotein cleavage signals. Co-detection analyses showed that each novel avian carbovirus was detected together with at least one additional virus in one or more datasets, although not consistently across all datasets. This recurrent but non-universal pattern raises the possibility of virus–virus associations, but their biological relevance remains to be tested. Together, these findings demonstrate the strength of *in silico* screening for revealing bornaviral diversity and underscore the need for experimental validation of predicted genomic features, host associations, and potential helper–virus relationships.

Acknowledgements

We thank the global scientific community for making raw sequencing data publicly available. Although these datasets were generated for specific research objectives, they contain additional valuable information that extends well beyond their original scope and enables further discoveries. The super-computing resources were in part provided by the Human Genome Center, the Institute of Medical Science, the University of Tokyo.

Supplementary material

Supplementary material is available at *VEVOLU Journal* online.

Conflicts of interest

The authors declare no conflicts of interest. The funders had no role in study design, data collection, analysis, interpretation, manuscript writing, or decision to publish the findings.

Funding

The investigations were supported by a Friedrich-Loeffler-Institut internal PhD program, grant number FLI-IVD-XX-2021-83 (F.P.) and JSPS KAKENHI grant numbers JP24K21922 (M.H.), JP22K1923 (M.H.), and JP21H01199 (M.H.), and the 2025 Osaka Metropolitan University (OMU) Strategic Research Promotion Project (Young Researcher) (M.K.).

Data availability

Nucleotide sequence data reported are available in DDBJ/ENA/GenBank databases under the accession numbers: BK075392-BK075421 and PZ005945. RNA-seq data from a stump-toed frog was submitted to SRA using the accession number SRR37117020 (PRJNA1418979).

References

Acchioni C, Sandini S, Acchioni M *et al.* Co-infections and superinfections between HIV-1 and other human viruses at the cellular level. *Pathogens* 2024;**13**. https://doi.org/10.3390/pathogens13050349

Bailey TL, Elkan C. 'Fitting a mixture model by expectation maximization to discover motifs in biopolymers'. *Proceedings International Conference on Intelligent Systems for Molecular Biology* 1994;**2**:28–36.

Briese T, La Torre JC d, Lewis A *et al.* 'Borna disease virus, a negative-strand RNA virus, transcribes in the nucleus of infected cells'. *Proc Natl Acad Sci* 1992;**89**:11486–9. https://doi.org/10.1073/pnas.89.23.11486

Briese T, Schneemann A, Lewis AJ *et al.* 'Genomic organization of Borna disease virus'. *Proc Natl Acad Sci* 1994;**91**:4362–6. https://doi.org/10.1073/pnas.91.10.4362

Broad Insitute. *Picard Toolkit*, 2019. https://broadinstitute.github.io/picard/

Buchfink B, Reuter K, Drost H-G. 'Sensitive protein alignments at tree-of-life scale using DIAMOND'. *Nat Methods* 2021;**18**:366–8. https://doi.org/10.1038/s41592-021-01101-x

Bushmanova E, Antipov D, Lapidus A *et al.* 'rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data'. *GigaScience* 2019;**8**. https://doi.org/10.1093/gigascience/giz100

Bushnell B, Rood J, Singer E. 'BBMerge - accurate paired shotgun read merging via overlap'. *PLoS One* 2017;**12**:e0185056. https://doi.org/10.1371/journal.pone.0185056

Chernomor O, Haeseler A v, Minh BQ. 'Terrace aware data structure for phylogenomic inference from Supermatrices'. *Syst Biol* 2016;**65**:997–1008. https://doi.org/10.1093/sysbio/syw037

Danecek P, Bonfield JK, Liddle J *et al.* 'Twelve years of SAMtools and BCFtools'. *GigaScience* 2021;**10**. https://doi.org/10.1093/gigascience/giab008

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 2007;**7**:214. https://doi.org/10.1186/1471-2148-7-214

Duckert P, Brunak S, Blom N. Prediction of proprotein convertase cleavage sites. *Protein Eng Des Sel* 2004;**17**:107–12. https://doi.org/10.1093/protein/gzh013

Edgar RC. 'MUSCLE: multiple sequence alignment with high accuracy and high throughput'. *Nucleic Acids Res* 2004;**32**:1792–7. https://doi.org/10.1093/nar/gkh340

Edgar RC, Taylor B, Lin V *et al.* 'Petabase-scale sequence alignment catalyses viral discovery'. *Nature* 2022;**602**:142–7. https://doi.org/10.1038/s41586-021-04332-2

Eshak MIY, Rubbenstroth D, Beer M *et al.* 'Diving deep into fish bornaviruses: uncovering hidden diversity and transcriptional strategies through comprehensive data mining'. *Evolution* 2023;**9**. https://doi.org/10.1093/ve/vead062

Eshak MIY, Breithaupt A, Tews BA *et al.* 'Wǔhàn sharpbelly bornavirus infects and persists in cypriniform cells'. *J Virol* 2025;**99**:e0132225. https://doi.org/10.1128/jvi.01322-25

Fujino K, Yamamoto Y, Daito T *et al.* 'Generation of a non-transmissive Borna disease virus vector lacking both matrix and glycoprotein genes'. *Microbiol Immunol* 2017;**61**:380–6. https://doi.org/10.1111/1348-0421.12505

Garry CE, Garry RF. Proteomics computational analyses suggest that the bornavirus glycoprotein is a class III viral fusion protein (gamma penetrene). *Virol J* 2009;**6**:145. https://doi.org/10.1186/1743-422X-6-145

González Aparicio LJ, López CB, Felt SA. 'A virus is a community: diversity within negative-sense RNA virus populations'. *Microbiol Mol Biol Rev* 2022;**86**:e0008621. https://doi.org/10.1128/mmbr.00086-21

Hallgren J, Tsirigos KD, Pedersen MD *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *bioRxiv* 2022. https://doi.org/10.1101/2022.04.08.487609

Hetzel U, Szirovicza L, Smura T *et al.* Identification of a novel Deltavirus in Boa constrictors. *MBio* 2019;**10**. https://doi.org/10.1128/mBio.00014-19

Hoang DT, Chernomor O, Haeseler A v *et al.* 'UFBoot2: improving the ultrafast bootstrap approximation'. *Mol Biol Evol* 2018;**35**:518–22. https://doi.org/10.1093/molbev/msx281

Hoffmann B, Tappe D, Höper D *et al.* 'A variegated squirrel Bornavirus associated with fatal human encephalitis'. *N Engl J Med* 2015;**373**:154–62. https://doi.org/10.1056/NEJMoa1415627

Hyndman TH, Shilton CM, Stenglein MD *et al.* 'Divergent bornaviruses from Australian carpet pythons with neurological disease date the origin of extant Bornaviridae prior to the end-cretaceous extinction'. *PLoS Pathog* 2018;**14**:e1006881. https://doi.org/10.1371/journal.ppat.1006881

Ji J, Li C, Hu T *et al.* 'Diverse RNA viruses in the venom-related microenvironment of different animal phyla'. *Virus Evol* 2024;**10**:veae024. https://doi.org/10.1093/ve/veae024

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* 'ModelFinder: fast model selection for accurate phylogenetic estimates'. *Nat Methods* 2017;**14**:587–9. https://doi.org/10.1038/nmeth.4285

Kanda T, Sakai M, Makino A *et al.* Exogenous expression of both matrix protein and glycoprotein facilitates infectious viral particle production of Borna disease virus 1'. *J Gen Virol* 2022;**103**. https://doi.org/10.1099/jgv.0.001767

Kolde, R. Pheatmap: pretty heatmaps. In CRAN: Contributed Packages. The R Foundation. 2010. https://doi.org/10.32614/cran.package.pheatmap

Krueger F. In: Krueger F (ed.), *Trim Galore*, 2019. https://doi.org/10.5281/zenodo.7598955

Kuhn JH, Dürrwald R, Bào Y *et al.* 'Taxonomic reorganization of the family Bornaviridae'. *Arch Virol* 2015;**160**:621–32. https://doi.org/10.1007/s00705-014-2276-z

Lempp FA, Urban S. 'Hepatitis Delta virus: replication strategy and upcoming therapeutic options for a neglected human pathogen'. *Viruses* 2017;**9**. https://doi.org/10.3390/v9070172

Lyra ML, Monteiro JPC, Rancilhac L *et al.* 'Initial Phylotranscriptomic confirmation of homoplastic evolution of the conspicuous coloration and bufoniform morphology of pumpkin-toadlets in the genus Brachycephalus'. *Toxins* 2021;**13**. https://doi.org/10.3390/toxins13110816

Martin M. 'Cutadapt removes adapter sequences from high-throughput sequencing reads'. *EMBnetjournal* 2011;**17**:10.

Minh BQ, Schmidt HA, Chernomor O *et al.* 'IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era'. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

Neumann P, Lieber D, Meyer S *et al.* 'Crystal structure of the Borna disease virus matrix protein (BDV-M) reveals ssRNA binding properties'. *Proc Natl Acad Sci* 2009;**106**:3710–5. https://doi.org/10.1073/pnas.0808101106

Neuwirth E. R Color Brewer: Color Brewer Palettes. 2022. https://CRAN.R-project.org/package=RColorBrewer

Niller HH, Angstwurm K, Rubbenstroth D *et al.* 'Zoonotic spillover infections with Borna disease virus 1 leading to fatal human encephalitis, 1999–2019: an epidemiological investigation'. *Lancet Infect Dis* 2020;**20**:467–77. https://doi.org/10.1016/S1473-3099(19)30546-8

Perez-Vargas J, Amirache F, Boson B *et al.* 'Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo'. *Nat Commun* 2019;**10**:2098. https://doi.org/10.1038/s41467-019-10117-z

Pfaff F, Rubbenstroth D. 'Two novel bornaviruses identified in colubrid and viperid snakes'. *Arch Virol* 2021;**166**:2611–4. https://doi.org/10.1007/s00705-021-05138-3

Pfaff F, Victoria RG, Linde J *et al. SRAminer*, 2023. https://doi.org/10.5281/ZENODO.8385126

Philadelpho NA, Rubbenstroth D, Guimarães MB *et al.* 'Survey of bornaviruses in pet psittacines in Brazil reveals a novel parrot bornavirus'. *Vet Microbiol* 2014;**174**:584–90. https://doi.org/10.1016/j.vetmic.2014.10.020

Poenisch M, Unterstab G, Wolff T *et al.* 'The X protein of Borna disease virus regulates viral polymerase activity through interaction with the P protein'. *J Gen Virol* 2004;**85**:1895–8. https://doi.org/10.1099/vir.0.80002-0

Poenisch M, Wille S, Ackermann A *et al.* 'The X protein of Borna disease virus serves essential functions in the viral multiplication cycle'. *J Virol* 2007;**81**:7297–9. https://doi.org/10.1128/JVI.02468-06

Poenisch M, Staeheli P, Schneider U. 'Viral accessory protein X stimulates the assembly of functional Borna disease virus polymerase complexes'. *J Gen Virol* 2008;**89**:1442–5. https://doi.org/10.1099/vir.0.2008/000638-0

Poenisch M, Burger N, Staeheli P *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**:4297–307. https://doi.org/10.1128/JVI.02321-08

Quinlan AR, Hall IM. 'BEDTools: a flexible suite of utilities for comparing genomic features'. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundaton ffor Statistial Computing, 2024.

Rakotoarison A, Scherz M, Glaw F *et al.* 'Describing the smaller majority: integrative taxonomy reveals twenty-six new species of tiny microhylid frogs (genus Stumpffia) from Madagascar'. *Vertebrate Zoology* 2017;**67**:271–398. https://doi.org/10.3897/vz.67.e31595

Rubbenstroth D. 'Avian Bornavirus research-a comprehensive review'. *Viruses* 2022;**14**. https://doi.org/10.3390/v14071513

Rubbenstroth D, Rinder M, Stein M *et al.* Avian bornaviruses are widely distributed in canary birds (Serinus canaria f. domestica). *Vet Microbiol* 2013;**165**:287–95. https://doi.org/10.1016/j.vetmic.2013.03.024

Rubbenstroth D, Schmidt V, Rinder M *et al.* 'Phylogenetic analysis supports horizontal transmission as a driving force of the spread of avian Bornaviruses'. *PLoS One* 2016;**11**:e0160936. https://doi.org/10.1371/journal.pone.0160936

Rubbenstroth D, Briese T, Dürrwald R *et al.* 'ICTV virus taxonomy profile: Bornaviridae'. *J Gen Virol* 2021;**102**. https://doi.org/10.1099/jgv.0.001613

Schneemann A, Schneider PA, Kim S *et al.* 'Identification of signal sequences that control transcription of Borna disease virus, a nonsegmented, negative-strand RNA virus'. *J Virol* 1994;**68**:6514–22. https://doi.org/10.1128/jvi.68.10.6514-6522.1994

Schneider PA, Schneemann A, Lipkin WI. 'RNA splicing in Borna disease virus, a nonsegmented, negative-strand RNA virus'. *J Virol* 1994;**68**:5007–12. https://doi.org/10.1128/jvi.68.8.5007-5012.1994

Sievers F, Wilm A, Dineen D *et al.* 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega'. *Mol Syst Biol* 2011;**7**:539. https://doi.org/10.1038/msb.2011.75

Suchard MA, Lemey P, Baele G *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10'. *Virus Evol* 2018;**4**:vey016. https://doi.org/10.1093/ve/vey016

Sureau C, Negro F. 'The hepatitis delta virus: replication and pathogenesis'. *J Hepatol* 2016;**64**:S102–16. https://doi.org/10.1016/j.jhep.2016.02.013

Szirovicza L, Hetzel U, Kipar A *et al.* 'Snake Deltavirus utilizes envelope proteins of different viruses to generate infectious particles'. *MBio* 2020;**11**.

Taylor JM. 'Hepatitis D virus replication'. *Cold Spring Harb Perspect Med* 2015;**5**. https://doi.org/10.1101/cshperspect.a021568

Tomonaga K, Kobayashi T, Lee BJ *et al.* 'Identification of alternative splicing and negative splicing activity of a nonsegmented negative-strand RNA virus, Borna disease virus'. *Proc Natl Acad Sci* 2000;**97**:12788–93. https://doi.org/10.1073/pnas.97.23.12788

Tomonaga K, Kobayashi T, Ikuta K. 'Molecular and cellular biology of Borna disease virus infection'. *Microbes Infect* 2002;**4**:491–500. https://doi.org/10.1016/S1286-4579(02)01564-2

Valieris R. *Parallel-Fastq-dump*: 2021. https://github.com/rvalieris/parallel-fastq-dump

Wernike K, Hoffmann B, Kalthoff D *et al.* Development and validation of a triplex real-time PCR assay for the rapid detection and differentiation of wild-type and glycoprotein E-deleted vaccine strains of bovine herpesvirus type 1'. *J Virol Methods* 2011;**174**:77–84. https://doi.org/10.1016/j.jviromet.2011.03.028

Wolff T, Pfleger R, Wehner T *et al.* 'A short leucine-rich sequence in the Borna disease virus p10 protein mediates association with the viral phospho- and nucleoproteins'. *J Gen Virol* 2000;**81**:939–47. https://doi.org/10.1099/0022-1317-81-4-939

Virus Evolution, 2026, **12**, veag033
https://doi.org/10.1093/ve/veag033
Research Article