PNAS
www.pnas.org

Supplementary Information for

Highly diversified shrew hepatitis B viruses corroborate ancient origins and divergent infection patterns of mammalian hepadnaviruses

Andrea Rasche, Felix Lehmann, Alexander König, Nora Goldmann, Victor M. Corman, Andres Moreira-Soto, Andreas Geipel, Debby van Riel, Yulia A. Vakulenko, Anna-Lena Sander, Hauke Niekamp, Ramona Kepper, Mathias Schlegel, Chantal Akoua-Koffi, Breno F. C. D. Souza, Foday Sahr, Ayodeji Olayemi, Vanessa Schulze, Rasa Petraityte-Burneikiene, Andris Kazaks, Kira A. A. T. Lowjaga, Joachim Geyer, Thijs Kuiken, Christian Drosten, Alexander N. Lukashev, Elisabeth Fichet-Calvet, Rainer G. Ulrich, Dieter Glebe*, Jan Felix Drexler*

Dieter Glebe
Email: Dieter.Glebe@viro.med.uni-giessen.de

Jan Felix Drexler
Email: felix.drexler@charite.de

This PDF file includes:

Contents

Supplementary Methods

Animal sampling

Permits for shrew sampling and transportation of samples from Africa to Germany were obtained from all African and German authorities. Permit IDs for sampling were: 35-9185.82/0261, 8.87-51.05.20.09.210, 7221.3-030/09, 22-2684-04-15-107/09 (Germany), 1155 MDCS/CAB-1/kss, LAROCS/2012/05/16, EN RU 25 8, GW/S/253 (Ivory Coast, Nigeria, Sierra Leone). In Germany, additional animals were collected by local forestry institutions as by-catches during vole monitoring as part of pest control measures. All efforts were made to minimize suffering of animals. Animals were captured using Sherman traps (Sherman Live Trap Co., Tallahassee, U.S.A) or snap trapping. Live-caught animals were euthanized with an overdose of fluoran as described (1). Morphological shrew genus identification was performed using common identification keys as described (2, 3).

Amplification of partial *cytochrome b* genes of host DNA was used for characterization on the species level (2). Shrew species distributions were retrieved from the IUCN Red List of Threatened Species (http://www.iucnredlist.org) and mapped using QGIS (www.qgis.org).

Nucleic acid purification

For nucleic acid purification approximately 30 mg of tissue was homogenized and extracted using the RNeasy Kit (Qiagen, Hilden, Germany) the MagNA Pure 96 DNA and Viral NA large-volume kit (Roche, Penzberg, Germany). Blood samples (10-50 µl volume) were purified using the MinElute kit (Qiagen) or DNA and Viral NA small-volume kit (Roche). Cell culture supernatants were purified with the High Pure Viral Nucleic Acid Kit (Roche).

Hepadnavirus detection, quantification and full-genome amplification

Hepadnavirus screening was performed using a broadly reactive hemi-nested PCR assay as described (4). Complete genome sequences and preC/Core partial sequences were acquired using strain-specific primers as described (4). Viral loads were quantified using strain-specific real-time PCR assays. The following oligonucleotides (5'-3') were used: HBV-GF-rtF, TCGGCGGCGTTTTATCATA; HBV-GF-rtP, FAM-TGCTCTTTCTCCTGCTGCTATGCCTCA-BHQ1; HBV-GF-rtR, CAGGTACCCCAACCAGCAA (CSHBV), and CIVHBV-rtF, CAAAGAGTGCAATGGCCAAGA and CIVHBV-rtP, FAM-TCGCAGTGCCAAACCTCCAGGCA-BHQ1; CIVHBV-rtR, TTGGAGGGCATGATGTTAGTGA ($MSHBV_{CIV}$). Photometrically quantified strain-specific plasmid standards containing the PCR target region cloned into pCR4-TOPO TA vectors (Thermo

Fisher Scientific, Waltham, U.S.A) were used for quantification. Thermocycling was done as described previously (4).

In cell culture, the purified RNA from supernatant was digested for 1 h at 37°C with RNase-Free DNase I (Thermo Fisher Scientific). Intracellular HDV RNA from infected cells was purified with the Nucleospin 8 RNA II Kit (Macherey Nagel, Düren, Germany) according to the manufacturer's instructions. The HDV genomes were quantified as described (4). Efficacy of HBV/CSHBV Gt A replication was assessed by quantifying newly produced viral DNA in the supernatant three days after transfection via a specific discriminating SYBR green quantitative PCR as described (5).

Evolutionary analyses

Phylogenetic analyses were performed with a dataset of selected primate HBV sequences as described (6) and all non-primate hepadnavirus sequences available on GenBank (https://www.ncbi.nlm.nih.gov/genbank/) on 15th December 2018. Operational taxonomic units (OTUs) were generated in CD-Hit using a cut-off of 95% sequence identity (7). Nucleotide and amino acid sequences were aligned in Geneious 9.1.8 (8) using MAFFT (9). Maximum likelihood (ML) and Neighbor Joining (NJ) phylogenetic trees were calculated in MEGA7 (10) using 1000 bootstrap replicates (11). A Hasegawa-Kishino-Yano (HKY) substitution model was used for ML phylogenies based on nucleotide sequences and a WAG substitution model for ML phylogenies based on amino acid sequences. Gaps and ambiguous data were deleted. Bayesian phylogenies of amino acid sequences were generated with MrBayes V3.1 (12). Ancestral state reconstructions (ASR) and hypothesis testing were generated using the BEAST package V1.8 (13) as described (6), a HKY+G substitution model and an uncorrelated relaxed clock with a lognormal distribution and a Yule speciation process as tree prior. Trees were midpoint rooted. The MCMC analysis was run for 50,000,000 generations sampled every 10,000th generation. Tree files were annotated using a burn-in of 25% in TreeAnnotator and visualized using FigTree (13). For hypothesis testing, MRCAs of monkey and shrew viruses were calibrated using priors describing the MRCAs of their hosts (normally distributed prior, 24.0 mya, 95% HPD, 22.6-25.3 for the Atelidae-Cebidae split and an exponentially distributed prior, 20 mya, 95% HPD, 19-25 for the Soricinae-Crocidurinae split) (14, 15). For calibration of the root of orthohepadnaviruses either a normally distributed prior describing the MRCA of the order Chiroptera (65.5 mya, 95% HPD, 64.8-66.2) or the clade Laurasiatheria (80.5 mya, 95% HPD, 79.7-81.3) was used (16). An uncorrelated relaxed lognormal clock calibrated according to Mühlemann et al., 2018 was used (17). To test which root prior was more compatible with the dataset, marginal likelihoods and associated Bayes factors were calculated using the stepping stone algorithm (18) with 50 path steps with chain lengths of 1,000,000 generations. Similarity scans were conducted in SSE with a window size of 250 nucleotides and a step

size of 25 nucleotides (19). Nucleotide pairwise distances were calculated in MEGA7 with a pairwise deletion option using MAFFT alignments generated in Geneious 9.1.8. Alignments included one virus sequence per OTU, resulting in a final 16 sequences for Chiroptera, 20 for Primates, 5 for Eulipotyphla and 3 for Rodentia. Subsequently Laurasiatheria and Euarchontoglires comprised 23 sequences each.

Signal peptidase cleavage sites were predicted using SignalP V4 (20). HBeAg homology modelling was done using Swiss-Prot (21) on human HBV HBeAg structures deposited under pdb accession file 6CVK. Structures were visualized and compared in Chimera (22). For clarity of presentation, only a shrew HBV monomer is overlaid on the human HBV HBeAg dimer. For modelling, the G1896A mutation was reverted in shrew HBV, leading to a tryptophan residue instead of a stop codon. RNA folding was predicted in mfold (23).

Recombination analyses

The alignment of complete hepadnavirus genomes (n=101) was subjected to recombination analysis using RDP4 software package (24) with default settings. Bootscanning was done using SimPlot 3.1 (25) using a window size of 200 nucleotides and a step size of 10 nucleotides. Recombination events were considered reliably supported if they were detected by at least two of the methods implemented in the RDP4 package. The initial scanning indicated recombination events among viruses infecting one host order (primates, or Old World bats, or shrews etc.). To exclude such events, multiple viruses infecting closely related hosts were omitted. In the remaining viruses (n=29), two genomic regions were devoid of reliably supported recombination breakpoints: alignment positions 3421-696 (counting started from the methionine of the SHBs ORF, this area encodes parts of polymerase and S proteins) and 1013-2209 (encoding X and partial core proteins). $MSHBV_{CHN}$ was only included in the fragment covering nt 3421-696, due to recombination with $MSHBV_{CIV}$ within the other fragment (**Fig. S1H**). To prove that there was no recombination between the two fragments, pairwise genetic distances in each alignment fragment were plotted against each other as described (26). The resulting plot (**Fig. S1I**) showed very good correspondence of genetic distances in two genome regions, which assumed absence of recombination between them; therefore, concatenating two non-recombinant fragments was deemed valid. Both genome regions were used for ASR and hypothesis testing as detailed above, both separately, and concatenated (**Fig. S1J**).

Virus abbreviations

HBV, hepatitis B virus; DCHBV, domestic cat HBV; MDHBV, Maxwell's duiker HBV; GSHV, ground squirrel hepatitis virus; ASHV, arctic squirrel hepatitis virus; WHV, woodchuck hepatitis virus; MSHBV, musk shrew HBV; CSHBV, crowned shrew HBV; GorHBV, gorilla HBV; ChimpHBV,

chimpanzee HBV; OrHBV, orangutan HBV; GibHBV, gibbon HBV; CMHBV, capuchin monkey HBV; WMHBV, woolly monkey HBV, RBHBV, roundleaf bat HBV; HBHBV, horseshoe bat HBV; PBHBV, pomona bat HBV; LBHBV, long-fingered bat HBV; TBHBV, tent-making bat HBV; BGHBV, bluegill HBV.

Rationale for virus designation

According to the International committee on Taxonomy of Viruses (ICTV) hepadnavirus species are termed based on common names of their host species. The African shrew hepadnavirus species comprised two divergent lineages from two different species both belonging to the genus *Crocidura* (musk shrew). This virus species was therefore termed based on the common name of the genus, musk shrew hepatitis B virus (MSHBV). The German shrew hepadnavirus also occurred in two species of the same genus (*Sorex*). However, both genotypes of the German shrew hepadnavirus species occurred almost exclusively in the species *S. coronatus* (crowned shrew). Only one out of 12 samples originated from *S. araneus* (common shrew). The genomic HBV sequence recovered from *S. araneus* was closely related to HBV genomes originating from *S. coronatus* (**Fig. S1A**), suggesting a spill-over from *S. coronatus* to *S. araneus*. Hence, this virus species is presumably mainly associated with *S. coronatus* and consequently termed crowned shrew hepatitis B virus (CSHBV). Sequences from Chinese shrews were also retrieved from different hosts, including the genera *Anourosorex* and *Crocidura*. Similar to CSHBV sequences, *Anourosorex* and *Crocidura* sequences were closely related, suggesting a spill-over from *Crocidura*-associated viruses given the *Crocidura*-associated monophyletic viruses from this study. We thus assume that the sequences retrieved from Chinese shrews are mainly *Crocidura*-associated and accordingly termed them musk shrew HBV. To distinguish both MSHBV species, a three-letter country code was added to the virus name, $MSHBV_{CHN}$ for the Chinese virus and $MSHBV_{CIV}$ for the African virus, since all complete genomes of the African musk shrew virus were acquired from Ivory Coast (Côte d'Ivoire).

Statistical analyses

Statistical tests were conducted in GraphPad Prism V7 (www.graphpad.com). For comparison of detection rates we used 2x2 contingency tables and two-tailed fisher exact test or chi-squared test with Yates' correction. For the comparison of viral concentrations, we conducted a two-tailed unpaired t-test comparing viral concentrations in liver to mean viral concentrations in blood and the combined other solid organs.

Ntcp sequencing and selection pressure analyses

Liver tissues were used for the characterization of taurocholate (Na+) cotransporting polypeptide (NTCP in humans, Ntcp in animals) coding mRNA sequences, using broadly reactive assays, specific primers and Rapid Amplification of cDNA Ends (RACE) strategies as described (6) and aligned with all Ntcp sequences from eulipotyphlan hosts available on GenBank, including the star-nosed mole (*Condylura cristata*), the European hedgehog (*Erinaceus europaeus*) and the common shrew (*Sorex araneus*). The final dataset encompassed GenBank accession numbers MN170823-MN170825, XM_004681741, XM_007535252, XM_004612415. Pressure analyses were done using the CodeML program implemented in the pamlX 1.3.1 software package (27, 28) as described (29) and the HyPhy package (Datamonkey.org) (30) using Fixed Effect Likelihood (FEL), Single Likelihood Ancestor Counting (SLAC), and Random Effect Likelihood (REL) (31), Mixed Effects Model of Evolution MEME (32), and Single Bayesian Approximation (FUBAR) (33). For all methods performed with the HyPhy package, an automatic model selection was run, and the proposed model with the best AIC (Akaike Information Criterion) was used.
The Eulipotyphla phylogeny (ML) was reconstructed in MEGA7 (10) using Ntcp sequences from all eulipotyphlan hosts and a bat Ntcp from *Rhinolophus sinicus* as an outgroup.

High-throughput sequencing

Library preparation and Illumina MiSeq sequencing was done using the KAPA Frag Kit, KAPA HyperPrep kit (Roche Molecular Diagnostics, Basel, Switzerland) and MiSeq reagent v2 chemistry (Illumina San Diego, U.S.A) according to the manufacturer's protocol. For single nucleotide polymorphism (SNP) analysis reads were trimmed and mapped against CSHBV/MSHBV$_{CIV}$ complete genome sequences using Geneious 9.1.8 (8). All SNPs occurring at a minimum coverage of 200 reads in a minimum variant frequency of 2% were analyzed. All reads with low quality were removed from the dataset. To exclude the occurrence of additional distinct hepadnaviruses in the samples, all reads were mapped against CSHBV and MSHBV$_{CIV}$ complete genomes and unmapped reads were blasted against an *Orthohepadnavirus* database.

In-situ hybridization

RNAScope RNA probes (Advanced Cell Diagnostics, Newark, U.S.A.) were custom designed based on the first 1,000 nucleotides following the start of the SHBs domain of CSHBV Gt A. *In situ* hybridization was performed as described by the manufacturer.

Cell lines and transfection

The human hepatoma cell lines HuH7 and HepG2 Tet-On (Clontech, Saint-Germain-en-Laye, France) were cultured in DMEM supplemented with 10% FCS. At a confluence of 80%, cells were transiently transfected with 200 ng plasmid per cm^2 cell growth area. For this, HuH7 were incubated with a mastermix using Fugene HD (Promega, Mannheim, Germany) and HepG2 cells with a mastermix using X-tremeGENE (Roche) according to the manufacturer's protocol. One day post transfection, cells were washed with PBS and medium was replaced with fresh DMEM supplemented with 2% FCS. The medium was changed every three days.

Peptide binding and taurocholate transport assay

HepG2 cells were plated at 60% confluency in DMEM supplemented with 10% FCS. The following day, cells were transfected with plasmids coding for respective NTCP/Ntcp under CMV promotor (34). After 48 h, cells were washed once with pre-warmed (37 °C) phenol red-free DMEM and incubated with respective myr-preS1 peptides at specified concentrations or 5 µM NBD-TC (fluorescent bile acid 4-nitro-2,1,3-benzoxadiazole taurocholic acid) in phenol red-free DMEM for 15 minutes at 37 °C (35). Cells were washed twice with phenol red-free DMEM and directly photographed using Leica fluorescence microscope DMI6000B (Leica, Wetzlar, Germany).

Production of hepatitis D virus pseudotype particles (psHDV)

To generate infectious psHDV, equimolar amounts of the expression plasmid pSVL(D3) encoding a trimer of the HDV genome and expression plasmids encoding hepadnaviral L-ORF were transfected with Fugene HD (Promega) into HuH7 cells according to the manufacturer's protocol. After transfection, cells were cultivated in DMEM supplemented with 2% FCS. Media were changed every three days. Secreted HDV was quantified by determining genome equivalents in supernatant by a specific reverse transcriptase quantitative PCR (RT-qPCR) on days 3, 6 and 9. The following oligonucleotides (5'-3') were used for HDV quantification: HDV-sense, TCCAGAGGACCCCTTCA; HDV-antisense; CCGGGATAAGCCTCACT; HDV-probe, FAM-AGACCGAAGCGAGGAGGAAAGCA-TAMRA. Supernatants from day 6 to day 9 were concentrated by membrane ultrafiltration (Vivaspin, MWCO 300 000; GE Healthcare, Solingen, Germany) according to the manufacturer's instructions and subsequently used for infection assays.

HDV pseudotype infection assays

HepG2 cells transiently expressing human or *Sorex* Ntcp were inoculated with similar genome equivalents (4 GE/cell) of HDV_{HBV}, $HDV_{CSHBV\ Gt\ A}$, $HDV_{MSHBV\text{-}CIV}$ in the presence of 4% PEG and 2% DMSO overnight. Sixteen hours after virus inoculation, cells were washed three times with hepatocyte growth medium (HGM) and fresh HGM was added. Medium was changed every three days until cells were fixed with 4% paraformaldehyde (20 min, RT) on day 9 post infection (p.i.), permeabilized with 0.2% Triton-X 100 in PBS (30 min, RT), blocked in 10% FCS (1 h, RT) and subsequently immunostained using human anti-HDV-positive serum (1 h, 37 °C) and anti-human-IgG antibody coupled to Alexa-568 (Invitrogen Karlsruhe, Germany) as secondary antibody (30 min, 37 °C).

Human primary hepatocytes (PHH) were obtained from Primacyte Cell Culture Technology (Schwerin, Germany). Primary hepatocytes from *Tupaia belangeri* (PTH) were obtained from an own breeding colony at the Justus Liebig University of Giessen. Permission to use primary hepatocytes from *Tupaia belangeri* was obtained from the Institutional Review Board of Justus Liebig University of Giessen and animals were handled according to ethical rules for animal care (JLU-No. 410M, Project-ID 767). Primary hepatocytes were inoculated with similar genome equivalents/cells (5 GE/cell) of HDV_{HBV} and $HDV_{CSHBV\ Gt\ A}$ in the presence of 4% PEG overnight. Sixteen hours after virus inoculation, cells were washed three times with hepatocyte growth medium (HGM) and fresh HGM was added. As infection specificity control, cells were pre- and co-incubated with different amounts of infection-interfering hepadnaviral myristoylated (myr)-preS1-peptides of amino acid residues 2-48 of HBV or CSHBV Gt A for 30 min, respectively. Twelve days p.i., total RNA of the cells was purified and HDV production was quantified using RT-qPCR as described above.

Immunofluorescence assays

For the detection of intracellular HBsAg expression, HuH7 cells were transiently transfected as described above with plasmids encoding the L-ORF of CSHBV Gt A or HBV, respectively. At day 4 post transfection, cells were washed twice with PBS and fixed using 4% paraformaldehyde (1 h, 4 °C). Cells were washed as described and permeabilized using 0.2% Triton-X 100 in PBS (30 min, RT). Subsequently, cells were washed again and blocked with 10% FCS in PBS (1 h, RT). After washing, cells were incubated with a dilution of polyclonal antisera (1 h, 37°C) from rabbits, which had been immunized with HBsAg of CSHBV Gt A or HBV, respectively (Eurogentec, Liège, Belgium). Then, cells were washed four times again and incubated with goat anti-rabbit-IgG secondary antibody coupled to Alexa Fluor-568 (Invitrogen) and DAPI (30 min, 37 °C). After washing four times with PBS, cells were stored at 4 °C in PBS until visualization by fluorescence microscopy.

For the detection of antibodies in blood samples from German shrews, a CSHBV Gt A 1.1 overlength construct was cloned into the mammalian pCEP expression vector via In-Fusion Cloning Kit (Takara, Saint-Germain-en-Laye, France) (4). This construct was used to transfect HuH7 cells as described above. At day 4 post transfection, cells were washed, fixed and permeabilized as described above. 86 blood samples originating from *S. coronatus* and *S. araneus* were diluted 1:20 in 0.1% FCS/DMEM. Cells were incubated with sample dilutions for 2 h at 37 °C and washed with PBS containing 0.1% Tween (PBS-T), three times for 5 min. Next, cells were incubated with a shrew specific antibody as described (36) (rabbit anti-suncus,1:200 in PBS) for 1 h at 37 °C, washed as described above and incubated for 30 min at 37 °C with a Cy3 labeled goat anti-rabbit antibody (1:100 in PBS) (Dianova, Hamburg, Germany). After final washing with PBS-T, cells were rinsed with water, dried and mounted with DAPI Gold Mounting Medium (Thermo Fisher Scientific).

For the detection of antibodies in blood samples from Sierra Leone, a full-length $MSHBV_{CIV}$ under endogenous viral promotors was cloned into the mammalian expression vector pcDNA3.1+. The vector was digested with MluI and BamHI (Thermo Fisher Scientific). Subsequently, two chemically synthesized subgenomic DNA fragments (IDT, Leuven, Belgium) totaling a 1.5 overlength genomic sequence were inserted into the linearized vector via In-Fusion recombination (Takara). Successful recombination was assessed via Sanger sequencing (LGC Genomics, Berlin, Germany). This construct was used for transfection of HepG2 Tet-On cells as described above. At day 4 post transfection, cells were washed and fixed as described. Immunofluorescence assay was conducted as detailed above, using 57 sera samples originating from *Crocidura*.

Replication inhibition with ETV

HuH7 cells were transfected with expression plasmids of the replication competent HBV and CSHBV Gt A 1.1 overlength construct (4). The day after transfection, the cells were divided into 6-well plates. Additionally, different concentrations of the reverse transcriptase inhibitor Entecavir (Sequoia Research Products, Pangbourne, UK) were added to the medium. After three days, replication efficiency was measured by quantifying newly produced viruses in the cell supernatant by a specific qPCR as described (5).

Generation of CSHBV surface particles in yeast

CSHBV SHBs encoding sequence was cloned into XbaI-linearized yeast expression plasmid pFX7 (37) and transformed into the yeast *S. cerevisiae* wild-type strain FH4C [ATCC 42368]. Transformation, cultivation of transformed yeast cells and protein synthesis induction, was performed as previously described (38, 39). After cultivation, the cells were harvested and suspended in a disruption buffer (PBS,

pH 7.4, 1 mM EDTA, 1 mM PMSF). The suspended cells were transferred to a chilled container for homogenization in a bead beater (BioSpec Products, Bartlesville, USA). Glass beads equivalent to the total weight of the cells and disruption buffer were added to the cells. The cells were homogenized in a bead beater for 5 min. After homogenization, the suspension was centrifuged at 2,000 ×g for 5 min at 4 ºC. The supernatant was collected and centrifuged at 10,000 ×g for 40 min at 4 ºC. Contaminating cellular proteins were extracted by suspending the pellets in the extraction buffer (PBS, pH 7.4; 1 mM EDTA; 1 mM PMSF, 0.5% Tween-20), and shaking for 30 min at 4 °C and centrifuged at 9,000 ×g at 4 °C for 20 min. The pellet suspended in PBS buffer were loaded onto a chilled 30% sucrose cushion and centrifuged at 100,000 ×g (Beckman Optima LE-80K Ultracentrifuge, Brea, USA) for 3 h at 4 °C. The supernatant was discarded, and the pellet was suspended in 1 ml of chilled PBS buffer. The suspension was then loaded on top of a chilled CsCl solution (2.5 g/mL) and centrifuged at 100,000 ×g for 24 h. Fractions of 1 ml were collected. Fractions containing protein were pooled, diluted with a fresh and chilled CsCl solution and re-centrifuged at 100,000 ×g for 24 h. Fractions containing recombinant protein were pooled and dialyzed against PBS. Purified CSHBV SHBs protein was placed on 300-mesh carbon coated palladium grids, negatively stained with 2% aqueous uranyl acetate solution and examined by Morgagni 268 electron microscope (FEI, Eindhoven, Netherlands).

Generation of polyclonal antisera

Rabbits were injected subcutaneously with 100 µg of CsCl-gradient purified subviral particles of the CSHBV SHBs protein containing Freund's adjuvant and given three booster injections after 2, 4 and 8 weeks followed by final bleeding (Eurogentec).

To generate antisera against HBV SHBs, rabbits were injected intramuscularly with human vaccine HBVAXPRO 40 (MSD Vaccins, Lyon, France), containing 40 µg of yeast-derived SHBs of HBV adsorbed to amorphous aluminium hydroxyphosphate sulfate as an adjuvant. Final bleeding was done after three booster injections at 2, 4 and 8 weeks (Eurogentec).

Supplementary Figures and Tables

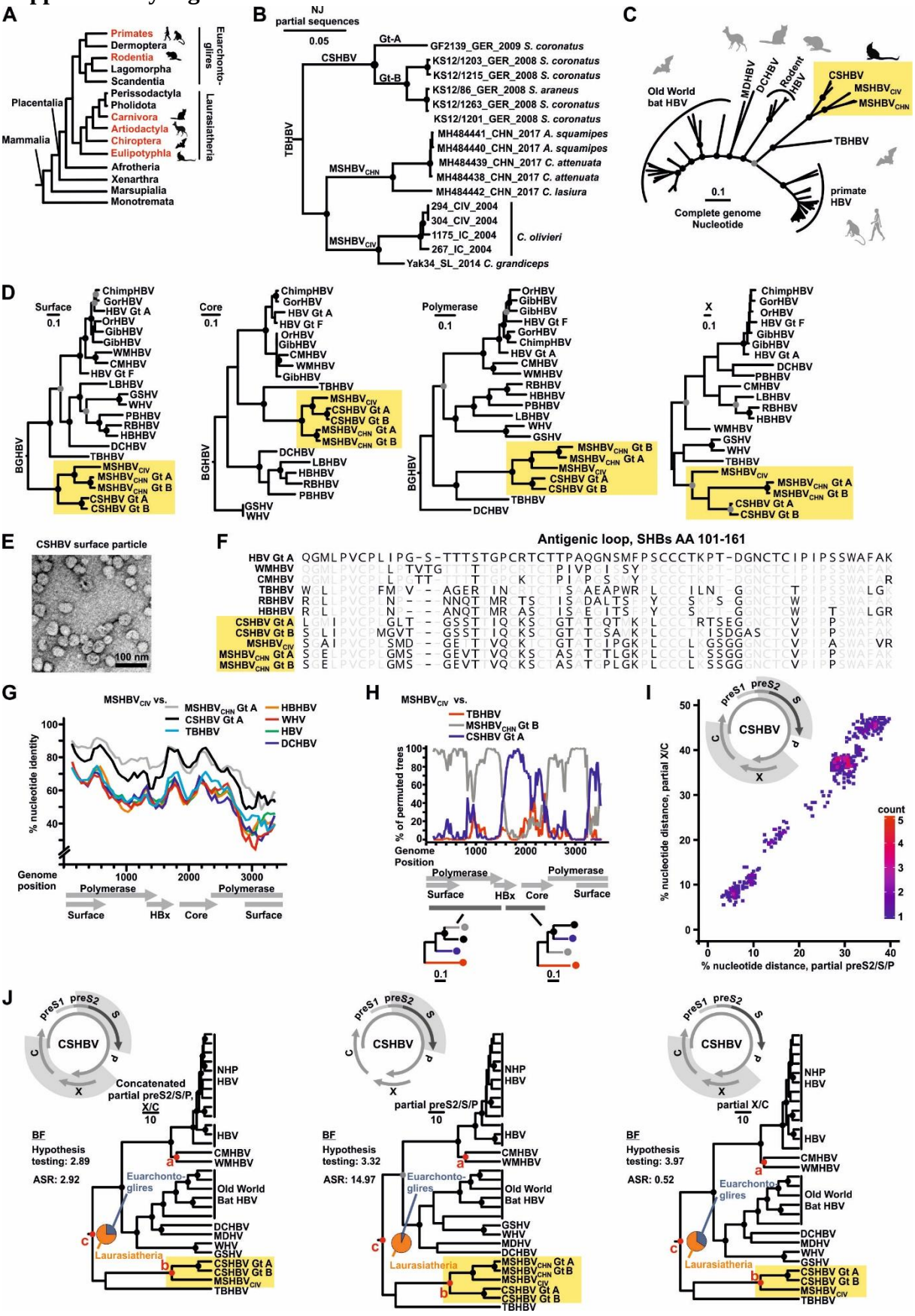

Fig. S1. Evolutionary analyses of shrew HBV

A) Cladogram of mammals. Red, host orders positive for HBV. Numbers at nodes, most recent common ancestors in million years (16). B) Phylogenetic analysis of partially characterized shrew hepadnaviruses Neighbor-Joining phylogeny using a partial sequence of 1,433 nucleotides encompassing partial SHBs, polymerase, HBx and core ORFs. C) Complete genome Maximum Likelihood (ML) phylogeny reconstructed in MEGA7 (10). Filled circles indicate bootstrap support of >95% (black) or >90% (grey). D) Bayesian phylogenies of translated surface, polymerase, and core open reading frames (ORFs) and ML phylogeny of the X ORF. E) Negative contrast electron microscopic images of CSHBV surface particles from yeast. F) Amino acid sequence identity of hepadnaviruses in the main antigenic region of the SHBs. G) Genomic identity between $MSHBV_{CIV}$ and reference sequences. Window size 250 nt, step size 25 nt. H) Bootscan graph of $MSHBV_{CIV}$ and reference sequences. Window size 200 nt, step 10 nt. Trees below the bootscan graph indicate reliably supported conflicting phylogenies in the genome regions indicated by grey bars. I) Pairwise nucleotide distance comparison plot. Each dot corresponds to a pair of nucleotide distances between the same pair of isolates in two genomic regions (axis labels): partial preS2/SHBs/Polymerase and partial X/Core. Color indicates the density of overlapping values. J) Bayesian ancestral state reconstruction and hypothesis testing (13) of partial sequences as highlighted in the genome. Pie charts, posterior probabilities for ancestral traits. Filled circles at nodes, posterior probability >0.95. Red nodes, priors used for hypothesis testing. Scale bar at phylogenetic trees indicate genetic distance. BF, Bayes Factor; ASR, Ancestral state reconstruction; *S.*, *Sorex*; *C.*, *Crocidura; A.*, *Anourosorex*; Ger, Germany; SL, Sierra Leone, IC, Ivory Coast; NHP, non-human primates. HBV, hepatitis B virus; CMHBV, capuchin monkey HBV; WMHBV, woolly monkey HBV; DCHBV, domestic cat HBV; MDHBV, Maxwell's duiker HBV; GSHV, ground squirrel hepatitis virus; WHV, woodchuck hepatitis virus; MSHBV, musk shrew HBV; CMHBV, crowned shrew HBV; TBHBV, tent-making bat HBV; RBHBV, roundleaf bat HBV; HBHBV, horseshoe bat HBV; Gt, genotype.

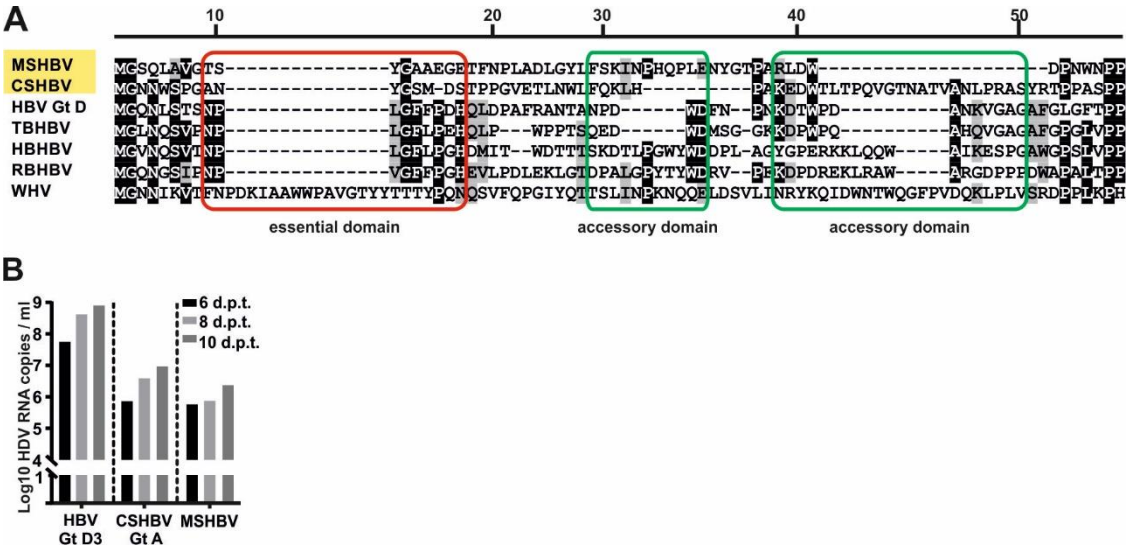

Fig. S2. N-terminal preS1 domain of selected hepadnaviruses and generation of pseudotyped HDV particles

A) Amino acid alignment of preS1 residues from selected orthohepadnaviruses highlighting domains interacting with NTCP/Ntcp. B) Generation of HDV particles pseudotyped with surface proteins of HBV, CSHBV and MSHBV$_{CIV}$ (HuH7 cells). MSHBV, musk shrew HBV; CMHBV, crowned shrew HBV; HBV, hepatitis B virus; TBHBV, tent-making bat HBV; HBHBV, horseshoe bat HBV; RBHBV, roundleaf bat HBV; WHV, woodchuck hepatitis virus; Gt, genotype.

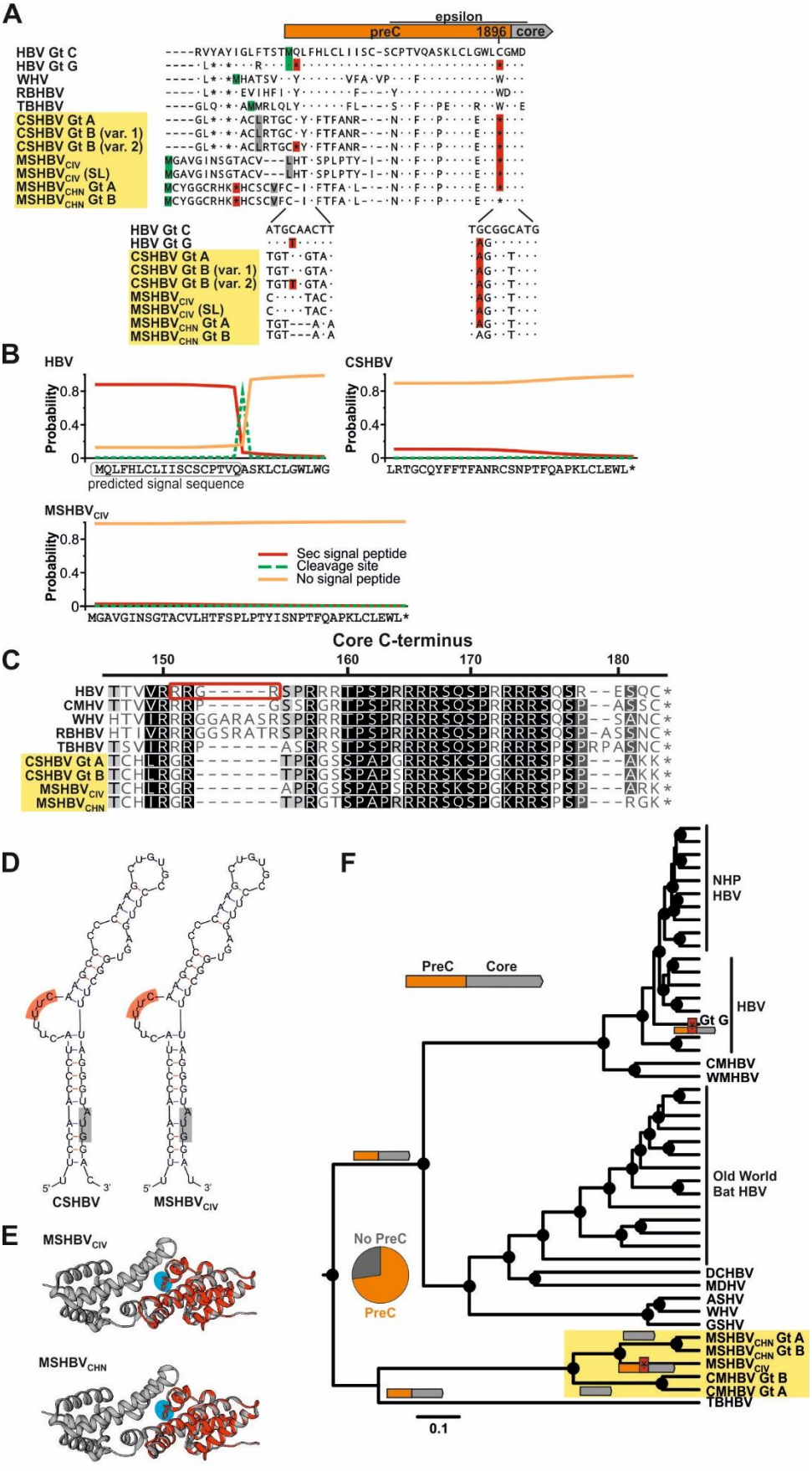
A
epsilon
preC
1896
core
HBV Gt C
HBV Gt G
WHV
RBHBV
TBHBV
CSHBV Gt A
CSHBV Gt B (var. 1)
CSHBV Gt B (var. 2)
MSHBV CIV
MSHBV CIV (SL)
MSHBV CHN Gt A
MSHBV CHN Gt B
B
HBV
CSHBV
Probability
predicted signal sequence
MSHBV CIV
Sec signal peptide
Cleavage site
No signal peptide
C
Core C-terminus
150
160
170
180
D
CSHBV
MSHBV CIV
E
MSHBV CIV
MSHBV CHN
F
PreC
Core
NHP
HBV
Gt G
CMHBV
WMHBV
Old World
Bat HBV
No PreC
PreC
DCHBV
MDHV
ASHV
WHV
GSHV
MSHBV CHN Gt A
MSHBV CHN Gt B
MSHBV CIV
CMHBV Gt B
CMHBV Gt A
TBHBV
0.1

Fig. S3. Lack of HBe antigen in shrew HBVs

A) Amino acid and partial nucleotide alignment of the preC-encoding region of selected reference hepadnaviruses and the homologous region of shrew hepadnaviruses. B) Signal peptide prediction of preC-encoding regions of HBV, CSHBV and $MSHBV_{CIV}$. C) Amino acid alignment of the C-terminus of the Core-encoding region. Predicted cleavage site in HBV is highlighted in red. D) RNA secondary structure predictions of the epsilon signal E) MSHBV HBeAg monomer (cyan, reverted G1896A) modeled on the HBV HBeAg dimer. F) Bayesian ancestral state reconstruction. Filled circles, posterior probability >0.95. Pie chart, posterior probability of occurrence and absence of a preC-encoding region. NHP, non-human primates. HBV, hepatitis B virus; CMHBV, capuchin monkey HBV; WMHBV, woolly monkey HBV; DCHBV, domestic cat HBV; MDHBV, Maxwell's duiker HBV; ASHV, arctic squirrel hepatitis virus; WHV, woodchuck hepatitis virus; MSHBV, musk shrew HBV; CMHBV, crowned shrew HBV; TBHBV, tent-making bat HBV; RBHBV, roundleaf bat HBV; HBHBV, horseshoe bat HBV; Gt, genotype.

Table S1. Sample characteristics

Country	Sampling years	Genus	Number of specimens	Number of PCR positive specimens (%), [95% CI]
Germany	2007-2014	*Sorex*	322	12 (3.7), [2.1-6.5]
Nigeria	2011-2013	*Crocidura*	146	0
Sierra Leone	2014-2016	*Crocidura*	134	1 (0.7), [<0.01-4.5]
Ivory Coast	2004-2005	*Crocidura*	91	4 (4.4), [1.4-11.1]
Total			693	17 (2.5), [1.5-3.9]

CI, confidence interval

Table S2. Genomic characteristics of selected orthohepadnaviruses

Species		Human (genotypes)								Monkey		Bat					Cat	Duiker	Rodent			Shrew			
		A	B	C	D	E	F	G	H	WMHBV	CMHBV	TBHBV	HBHBV	RBHBV	LBHBV	PBHBV	DCHBV	MDHBV	GSHV	ASHV	WHV	CSHBV (geno-types)		$MSHBV_{CIV}$	$MSHBV_{CHN}$
																						A	B		
Genome length (nt)		3221	3215	3215	3182	3212	3215	3248	3215	3179	3182	3149	3377	3368	3230	3278	3187	3128	3311	3302	3308	3089	3089	3172	3165
Viral proteins (amino acids)																									
Core		185	183	183	183	183	183	195	183	182	183	188	189	189	189	189	190	192	187	187	188	181	181	181	180
PreC		29	29	29	29	29	29	*	29	29	29	33	28	28	28	28	28	28	30	30	30	ø	ø	*	ø
Surface proteins	**LHBs**	400	400	400	389	399	400	399	400	391	390	382	448	445	399	415	382	362	428	427	426	367	367	395	395
	SHBs	226	226	226	226	226	226	226	226	228	227	223	224	224	224	224	226	225	222	222	222	225	225	225	225
	PreS2	55	55	55	55	55	55	55	55	55	55	50	47	47	45	47	52	35[§]	60	60	60	65[§]	65[§]	104[§]	53[§]
	PreS1	119	119	119	108	118	119	118	119	108	108	109	177	174	130	144	104	102	146	145	144	77	77	66	117
Polymerase		845	843	843	832	842	843	842	843	835	833	827	902	899	853	868	838	819	881	877	879	812	823	835	834
HBx		155	155	155	155	155	155	155	155	152	153	135	141	141	144	142	145	144	138	138	141	135	135	136	135

*stop codon terminates preC open reading frame

[§]predicted sequence, unusually long and/or no typical N-linked glycosylation signal: existence unclear

Table S3. Genomic distances between shrew HBV and other hepadnavirus species

Percentage nucleotide distance, complete genome					
	CSHBV Gt A	**CSHBV Gt B**	**$MSHBV_{CIV}$**	**$MSHBV_{CHN}$ Gt A**	**$MSHBV_{CHN}$ Gt B**
CSHBV Gt B	11.6-12.0				
$MSHBV_{CIV}$	26.5-26.8	26.9-27.4			
$MSHBV_{CHN}$ Gt A	28.5-28.7	28.5-28.9	22.9-23.1		
$MSHBV_{CHN}$ Gt B	27.9	27.6-28.0	23.1-23.3	8.5-8.7	
TBHBV	39.0	39.1-39.3	40.8-40.9	40.7-40.8	40.3
HBV	42.0	41.7-42.1	42.5-42.8	42.4	42.2
CMHBV	41.7	41.9-42.2	42.6-42.9	42.5-42.6	42.5
WMHBV	42.2	42.2-43.0	42.3-42.4	42.3-42.5	41.7
WHV	42.2	42.2-42.8	42.9	42.5-42.6	42.5
GSHBV	41.9	41.2-42.0	42.2-42.4	42.1-42.2	41.9
DCHBV	43.2	43.2-43.4	42.2-42.6	43.8	43.8
HBHBV	42.1	41.4-41.7	42.9-43.2	43.2-43.3	42.1
RBHBV	41.6	41.4-41.7	43.1-43.5	43.8-43.9	43.0
PBHBV	41.9	42.0-42.1	42.1	43.6-43.7	42.9
LBHBV	42.9	42.2-42.4	42.3-42.4	42.3	41.9
DHBV	57.4	57.0-57.2	56.8-57.1	57.8-57.9	58.4

Table S4. Selection pressure analyses of eulipotyphlan Ntcp

AA[1]	SLAC	FEL	REL	FUBAR	PAML[2]	MEME
L10	n. s.	n. s.	n. s.	n. s.	n. s.	0.0180
L30	n. s.	n. s.	n. s.	n. s.	n. s.	0.014
K157	**n. s.**	**n. s.**	**n. s.**	**n. s.**	**0.986**	**n. s.**
S213	n. s.	n. s.	n. s.	n. s.	n. s.	0.0156
T347	n. s.	0.095	92	0.929	0.998	0.0554
Q349	n. s.	n. s.	404	n. s.	0.999	n. s.
P350	n. s.	0.068	n. s.	n. s.	n. s.	0.0858
T352	n. s.	n. s.	58	n. s.	0.995	n. s.
G353	n. s.	n. s.	104	n. s.	0.999	n. s.

Statistical criteria: SLAC, FEL, MEME $p<0.1$, REL Bayes Factor>50, FUBAR and M8 BEB PPr>0.9.
[1]AA position in S*orex coronatus* Ntcp
[2]M8 BEB
Bold, within NTCP binding site

Table S5. Genomic variability within partial shrew HBV genomes

		SNPs per 100 nucleotides				
Sample ID	**Host species**	**Amplicon 1 surface/pol**	**Amplicon 2 X/preC/core**	**Double-coding**	**Single-coding**	**Non-coding**
KS12/86	*Sorex araneus*	1.11	0.48	0.97	0.38	0.99
KS12/1201	*Sorex coronatus*	0	0.72	0.48	0	0.99
KS12/1203	*Sorex coronatus*	0.28	0.24	0.24	0.38	0
KS12/1215	*Sorex coronatus*	2.79	1.68	2.66	1.92	0.99
KS12/1217	*Sorex coronatus*	0.28	0.24	0.24	0.38	0
KS12/1230	*Sorex coronatus*	0	0.24	0	0.38	0
KS12/1249	*Sorex coronatus*	0.56	0	0.48	0	0
KS12/1263	*Sorex coronatus*	0	0	0	0	0
KS12/1264	*Sorex coronatus*	0.28	n.d.	0.28	n.d.	n.d.
KS12/1265	*Sorex coronatus*	0	0.72	0.24	0.38	0.99
CIV267	*Crocidura olivieri*	0.56	0.74	0.54	0.68	1
CIV294	*Crocidura olivieri*	0.28	0.50	0.27	0.68	0
CIV304	*Crocidura olivieri*	0.28	0.50	0.27	0.68	0
CIV1175	*Crocidura olivieri*	0.56	0.50	0.54	0.68	0
Yak34	*Crocidura grandiceps*	0.56	0.50	0.54	0.34	1

HBV, hepatitis B virus; SNP, single nucleotide polymorphism; pol, polymerase; n.d., not determined

References

1. Johnson N (2014) The Role of Animals in Emerging Viral Diseases Foreword. *Role of Animals in Emerging Viral Diseases*:Xv-Xvii.
2. Schlegel M, *et al.* (2012) Molecular identification of small mammal species using novel cytochrome B gene-derived degenerated primers. *Biochemical genetics* 50(5-6):440-447.
3. Happold DCD (2013) *Mammals of Africa* (Bloomsbury Publishing, London).
4. Drexler JF, *et al.* (2013) Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc Natl Acad Sci U S A* 110(40):16151-16156.
5. Geipel A, *et al.* (2015) Entecavir allows an unexpectedly high residual replication of HBV mutants resistant to lamivudine. *Antiviral therapy* 20(8):779-787.
6. de Carvalho Dominguez Souza BF, *et al.* (2018) A novel hepatitis B virus species discovered in capuchin monkeys sheds new light on the evolution of primate hepadnaviruses. *J Hepatol* 68(6):1114-1122.
7. Huang Y, Niu B, Gao Y, Fu L, & Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* 26(5):680-682.
8. Kearse M, *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647-1649.
9. Katoh K, Misawa K, Kuma K, & Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 30(14):3059-3066.
10. Kumar S, Stecher G, & Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 33(7):1870-1874.
11. Krogh A, Larsson B, von Heijne G, & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3):567-580.
12. Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* 19(12):1572-1574.
13. Drummond AJ, Suchard MA, Xie D, & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969-1973.
14. Springer MS, *et al.* (2012) Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE* 7(11):e49521.
15. He K, *et al.* (2010) A multi-locus phylogeny of Nectogalini shrews and influences of the paleoclimate on speciation and evolution. *Molecular phylogenetics and evolution* 56(2):734-746.
16. Foley NM, Springer MS, & Teeling EC (2016) Mammal madness: is the mammal tree of life not yet resolved? *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 371(1699).
17. Muhlemann B, *et al.* (2018) Ancient hepatitis B viruses from the Bronze Age to the Medieval period. *Nature* 557(7705):418-423.
18. Baele G, Lemey P, & Vansteelandt S (2013) Make the most of your samples: Bayes factor estimators for high-dimensional models of sequence evolution. *BMC Bioinformatics* 14:85.
19. Simmonds P (2012) SSE: a nucleotide and amino acid sequence analysis platform. *BMC research notes* 5:50.
20. Petersen TN, Brunak S, von Heijne G, & Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* 8(10):785-786.
21. Waterhouse A, *et al.* (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* 46(W1):W296-W303.
22. Pettersen EF, *et al.* (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25(13):1605-1612.
23. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406-3415.

24. Martin DP, Murrell B, Golden M, Khoosal A, & Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus evolution* 1(1):vev003.
25. Lole KS, *et al.* (1999) Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73(1):152-160.
26. Deviatkin AA & Lukashev AN (2018) Recombination in the rabies virus and other lyssaviruses. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 60:97-102.
27. Xu B & Yang Z (2013) PAMLX: a graphical user interface for PAML. *Molecular biology and evolution* 30(12):2723-2724.
28. Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences : CABIOS* 13(5):555-556.
29. Feng H, *et al.* (2019) Hepatovirus 3ABC proteases and evolution of mitochondrial antiviral signaling protein (MAVS). *J Hepatol* 71(1):25-34.
30. Weaver S, *et al.* (2018) Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Molecular biology and evolution*.
31. Kosakovsky Pond SL & Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and evolution* 22(5):1208-1222.
32. Murrell B, *et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS genetics* 8(7):e1002764.
33. Murrell B, *et al.* (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular biology and evolution* 30(5):1196-1205.
34. Konig A, *et al.* (2014) Kinetics of the bile acid transporter and hepatitis B virus receptor Na+/taurocholate cotransporting polypeptide (NTCP) in hepatocytes. *J Hepatol* 61(4):867-875.
35. Petzinger E, Wickboldt A, Pagels P, Starke D, & Kramer W (1999) Hepatobiliary transport of bile acid amino acid, bile acid peptide, and bile acid oligonucleotide conjugates in rats. *Hepatology* 30(5):1257-1268.
36. Schlegel M, *et al.* (2012) Novel serological tools for detection of Thottapalayam virus, a Soricomorpha-borne hantavirus. *Archives of virology* 157(11):2179-2187.
37. Sasnauskas K, *et al.* (1999) Yeast cells allow high-level expression and formation of polyomavirus-like particles. *Biological chemistry* 380(3):381-386.
38. Samuel D, *et al.* (2002) High level expression of recombinant mumps nucleoprotein in Saccharomyces cerevisiae and its evaluation in mumps IgM serology. *J Med Virol* 66(1):123-130.
39. Samuel D, *et al.* (2003) Development of a measles specific IgM ELISA for use with serum and oral fluid samples using recombinant measles nucleoprotein produced in Saccharomyces cerevisiae. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 28(2):121-129.