PNAS
www.pnas.org

Supplementary Information for

Diversification of mammalian deltaviruses by host shifting

Laura M. Bergner, Richard J. Orton, Alice Broos, Carlos Tello, Daniel J. Becker, Jorge E. Carrera, Arvind H. Patel, Roman Biek, Daniel G. Streicker

*Corresponding authors: Laura Bergner (Laura.Bergner@glasgow.ac.uk) and Daniel Streicker (Daniel.Streicker@glasgow.ac.uk)

This PDF file includes:

Supplementary Results
Figures S1-S9
Table S1-S6
Dataset S1-S4 Descriptions
Supplementary References

Other supplementary materials for this manuscript include the following:

Dataset S1-S4

www.pnas.org/cgi/doi/10.1073/pnas.2019907118

Supplementary Results

1. Evaluation of continent level geographic biases in Serratus data

We sought to confirm that the sole detection of mammalian deltaviruses in the Americas in three different mammalian orders was unlikely to have arisen from sampling biases in the RNA sequence datasets that we analyzed. The main text illustrates geographic patterns at broad scales and shows that New World species were under-represented relative to Old World species, indicating that sampling biases were unlikely to explain the absence of deltaviruses from Old World species. However, we also examined biases at finer geographic and taxonomic scales, focusing on the mammalian SRA search query dataset for which there was species and continent level information. At the continent level, data volumes (in bases of RNA sequenced) declined from Asia (6.35e13), Africa (2.89e13), South America (7.99e12), North America (4.93e12), Europe (4.86e12), to Australia (3.75e12) (Fig S8). This implies that Africa (the previously assumed origin of HDV) has more than double the data of the Americas combined. Similar patterns were evident within the deltavirus-infected mammalian orders. For Artiodactyla, North and South American datasets were ranked 3rd and 5th respectively among the 5 continents which had sequence data. Although there was less Artiodactyla data from Africa than North America, Asia (1st) and Europe (2nd) had 2 and 1.6-fold more data than the Americas combined. For bats, North and South American datasets were ranked 5th and 6th among the 6 continents with data, and Europe (1st) and Africa (2nd) had 2.5 and 1.2-fold more data than datasets combined across the Americas. For rodents, North America datasets were ranked 2nd, while South American datasets were ranked 5th out of the 5 continents with data, but Asia (1st) had 1.1-fold more data than the Americas combined and Africa had 3.6 times more data than South America.

We also examined potential biases in search effort by number of species sequenced per continent. There were equal or fewer species of Artiodactyla sequenced in North (N = 5) and South America (N=1) compared to Europe (N = 5), Asia (N = 8) and Africa (N = 14). Similarly, rodent species sampled from Old World continents (N = 48; Europe [N = 8], Africa [N = 14], and Asia [N = 26]) outnumbered those in the New World (N = 33; North America [N = 21], South America [N = 12]. Neither Rodentia nor Artiodactyla datasets were searched from Australia. In contrast, there were more bat species in North and South American datasets (N = 4 and 42, respectively) compared to Asia (N = 22), Africa (N = 5), Europe (N = 2), and Australia (N =1) leading to a slight bias toward New World bats (46 species vs 30 Old World species). Consequently, in the ‘mammalian’ dataset, New World bats were more numerous at the species level, but had fewer individuals tested per species and/or less sequencing depth per species.

We further examined datasets generated by related search queries (vertebrate, metagenome and virome) which included some libraries from mammals that were excluded from the ‘mammalian’ dataset (see Materials and Methods, Section 1e). The vertebrate dataset contained no matches to the three mammalian Orders of interest (Rodentia, Artiodactyla,

Chiroptera). For Rodentia, the metagenome dataset contained 380 libraries (9.7e7 bases) identified as “mouse metagenome” and the virome dataset contained 171 libraries (1e11 bases) which were identified as “mouse gut metagenome” or “rodent metagenome” or “Rattus” or “rat gut metagenome”. These could not be assigned geographic provenance and likely represented laboratory animals. For Artiodactyla, the virome dataset contained 317 libraries (1e11 bases) which were identified as “pig gut metagenome” or “pig metagenome” or “bovine gut metagenome”. As these libraries likely derived from either experimental or domestic animals, we conclude that additional data from these two Orders are unlikely to influence geographic or taxonomic bias. For Chiroptera, there were a total of 25 libraries (2.4e10 bases) identified as “bat metagenome”. The vast majority (N=24) were from Old World bats. Eleven derived from a study of bat rotaviruses (PRJNA562472) which of which 10 were collected from Old world locations (Ghana, Bulgaria) and one from New World (Costa Rica). Two libraries were bat viral metagenomes generated from samples collected in South Africa (SRR5889194; SRR5889129), and twelve libraries were generated from bats sampled in China (PRJNA379515). There were 14 bat species analyzed in all of these libraries, of which eight were not included in the mammalian SRA dataset, bringing the total number of Old World bat species to 38. Therefore, by the metric of number of species, New World bats remained slightly over-represented (38 versus 46 species) though as mentioned above, Old World bat derived datasets were sequenced more comprehensively and covered a larger number of continents.

Overall, across the three orders in which we detected deltaviruses, fewer species were studied in North and South American datasets (85 species) compared to those from Africa, Asia, Europe and Australia (105 species) and the total volume of RNA sequenced was 2.7 times greater for Old world species (1.64e13 bases RNA) than New world species (6.02e12). We therefore conclude that the exclusive presence of deltaviruses in American mammals is unlikely to represent geographic biases in our datasets.

2. Large delta antigen in novel mammalian deltaviruses

In HDV, the large delta antigen protein (L-HDAg) is produced by RNA editing of the UAG stop codon to include 19 additional aa (1) and contains a farnesylation site which interacts with HBV (2). The DrDV-B DAg from the genome from bat colony CAJ1 terminated in UAG, which if edited similarly to HDV would generate a putative L-DAg containing an additional 28 aa (Fig. S3). In contrast, DrDV-B DAg from the two other bat colonies from which genomes were sequenced (LMA6 and AYA11), as well as DrDV-A DAg, terminated in a UAA stop codon so would not appear to be similarly edited, although it is possible to extend the open reading frames through frameshifting (3). Importantly, no putative vampire bat L-DAg generated through either RNA editing or frameshifting contained a farnesylation site. PmacDV, OvirDV, and MmonDV also did not contain apparent L-DAg extensions or farnesylation sites (4).

3. Co-phylogenetic results across posterior distributions of trees from two Bayesian searches and two different co-phylogeny analyses

Consensus topologies differed slightly between the phylogenetic analyses of deltaviruses using a multi-species coalescent model in StarBeast and a coalescent model in MrBayes, particularly in relation to the termite-associated deltavirus-like agent and avian deltavirus (compare Fig. 2A and 2B). We therefore repeated our co-phylogenetic analysis using 1,000 trees from the MrBayes analysis to verify that our conclusions were robust to this topological inconsistency. Analyses performed using virus distance matrices derived from posterior MrBayes trees were congruent with those in StarBeast.

In the broadest analyses of all taxa, results differed slightly among co-phylogenetic analyses. Specifically, PACo analyses supported the dependence of the deltavirus phylogeny on the host phylogeny (StarBeast trees: m^2_{xy} = 0.57 (standard deviation = 0.48- 0.66); $m^2_{xy\_null}$ = 1.27 (1.11- 1.44); *P* = 0.01; MrBayes trees: (m^2_{xy} = 0.53 (0.46- 0.6); $m^2_{xy\_null}$ = 1.26 (1.1- 1.42); *P* = 0.002). However, ParaFit analyses supported independence of the virus and host phylogenies based on both StarBeast trees (ParaFitGlobal = 0.72 (standard deviation = 0.59- 0.86); *P*= 0.09) and MrBayes trees (ParaFitGlobal = 0.63 (0.52- 0.75); *P*= 0.07). Similarly, for ingroup taxa (mammalian, avian and snake deltaviruses), PACo detected evidence of co-phylogeny using both sets of trees (Starbeast: m^2_{xy} = 0.81 (0.66- 0.96); $m^2_{xy\_null}$ = 1.3 (1.14- 1.46); *P* = 0.02; MrBayes: m^2_{xy} = 0.76 (0.69- 0.83); $m^2_{xy\_null}$ = 1.28 (1.1- 1.46); *P* = 0.01), while ParaFit analyses with StarBeast trees (ParaFitGlobal = 0.8 (0.7- 0.89); *P*= 0.12) and MrBayes trees (ParaFitGlobal = 0.68 (0.62- 0.75); *P*= 0.09) found no significant support for co-phylogeny.

All analyses of mammalian deltaviruses failed to reject the null hypothesis of independence of phylogenies. These results were consistent when with both StarBeast and MrBayes trees in PACo (StarBeast: m^2_{xy} = 1.51 (1.36- 1.65); $m^2_{xy\_null}$ = 1.66 (1.43- 1.9); *P* = 0.28; MrBayes: m^2_{xy} = 1.16 (1.07- 1.25); $m^2_{xy\_null}$ = 1.22 (0.97- 1.48); *P* = 0.35), as well as in ParaFit (StarBeast: ParaFitGlobal = 0.61 (0.55- 0.68); *P*= 0.52; MrBayes: ParaFitGlobal = 0.49 (0.44- 0.54); *P*= 0.5).

In summary, both PACo and ParaFit analyses of StarBeast and MrBayes trees showed no support for co-phylogeny in the mammalian dataset. Including more divergent host-virus pairs increased support for co-phylogeny in the all taxa and ingroup datasets, with these results being statistically significant in PACo analyses but not in ParaFit analyses. Inconsistent support for phylogenetic independence at broader scales may reflect variation in the sensitivity of different analyses to detect phylogenetic congruence which occurs in only a subset of branches. For example, the non-ingroup deltavirus-like agents formed a polytomy of long branches and were found in the most divergent hosts from mammals, which may have inflated co-phylogenetic signal (Fig. S5). Regardless, given the consistent evidence against a co-speciation model among mammals and incongruences observed among other taxa in the consensus topologies, these

findings illustrate that a model of co-speciation alone cannot explain the evolutionary relationships of deltaviruses and their hosts.

4. Putative cross-species transmission of DrDV-B to a frugivorous bat.

The detection of a vampire bat associated deltavirus in a frugivorous bat (*Carollia perspicillata)* is strongly suggestive of cross-species transmission but might also arise through mis-assignment of bat species in the field or contamination of samples during laboratory processing. To exclude the possibility of host species mis-identification, we confirmed morphological species assignment by sequencing Cytochrome B from the same saliva sample in which we amplified deltavirus (see Methods), which showed 99.49% identity with a published *C. perspicillata* sequence in Genbank (Accession AF511977.1). Laboratory contamination was minimized by processing all samples through a dedicated PCR pipeline with a one directional workflow. PCR reagents are stored and master mixes prepared in a laboratory that is DNA/RNA free, and which cannot be entered after going into any other lab. Field collected samples from bats are extracted and handled in a room strictly used for clinical samples which cannot be entered after going in any other lab aside from the master mix room. To further exclude laboratory contamination, we independently amplified the *C. perspicillata* deltavirus product from two separate batches of cDNA. We used only round 1 primers of a nested PCR to avoid detecting trace amounts of potential contamination; in vampire bats only 68% of individuals deemed positive after round 2 were also positive in round 1. Furthermore, in the laboratory, samples from other bat species were handled separately from samples collected from vampire bats, with extractions and PCRs being performed on different days. As discussed in the main text, the absence of genetic divergence from sympatric strains in *D. rotundus* indicates limited or no onward transmission of DrDV-B in *C. perspicillata*. Whether the *C. perspicillata* sustained an actively replicating infection is uncertain, although detection in a single round of PCR (which was true for only 68% of DrDV-positive vampire bats) implies an intensity of infection which could suggest DrDV replication in the recipient host, though this would require further testing to confirm. Definitively resolving the extent of DrDV-B replication could be achieved using a quantitative RT-PCR targeting the DrDV antigenome. Such assays do not currently exist and after the confirmatory testing above, in addition to metagenomic sequencing, we unfortunately would no longer have sufficient RNA available from the *C. perspicillata* bat to run such a test if it were available. In summary, we are confident that the individual in which the deltavirus was detected is a *C. perspicillata* and we believe the most likely explanation to be cross-species transmission in nature, though whether this represents an active infection remains uncertain.

5. Candidate helper viruses of OvirDV and MmonDV

We also examined viral communities in *O. virginianus* and *M. monax* libraries for candidate helpers. Given that MmonDV was detected in animals experimentally inoculated with *Woodchuck hepatitis virus* (WHV, a hepadnavirus), these libraries were unsurprisingly dominated by WHV,

but also contained reads matching to *Herpesviridae*, *Flaviviridae*, *Poxviridae* and *Retroviridae* (Figure S9). *O. virginianus* libraries contained *Poxviridae*, *Retroviridae*, and *Herpesviridae* reads. Consequently, reads matching to *Poxviridae* were detected in libraries for all deltavirus hosts which were studied here (DrDV-A, DrDV-B, PmacDV, MmonDV, OvirDV), although reads were less abundant than other viral taxa and could not always be decisively ruled out as false positives. Indeed, blastn analysis of poxvirus-like reads, which were originally identified by blastx, revealed that these reads frequently had poor correspondence to *Poxviridae* at the nucleotide level. Although there is no experimental evidence that poxviruses can produce infectious deltavirus particles, this putative ecological association may be worth considering in future studies of mammalian deltaviruses.

Supplementary Figures

	U81989	AB118846	AB118845	AM183326	AM183329	AM183333	AM183327	AB037948	OvirDV	MmonDV	DrDV-A	PmacDV	MK598009	DrDV-B	NC_040729	NC_040845	MN031240	MN031239	MK962760	MK962759
U81989_HDV1		73.5	72.3	69.6	68.3	68.9	71.2	60.4	57.8	51.5	53	46.8	46.8	47.5	41.8	37.2	31.9	30.7	28.1	30.1
AB118846_HDV2	73		76	74.7	74.1	71.5	72.5	62.5	58.7	52.6	54.5	47.5	47.4	48.2	42	37.5	33.2	31.7	28.5	30.2
AB118845_HDV4	74.4	75.7		74.3	72.6	71.5	72.9	63.1	61.5	53.4	56.1	48	48	49	41.8	37.4	32.9	30.4	27.8	29.9
AM183326_HDV5	70.9	77.1	75.3		74.7	72.6	72.1	61	58.6	51.2	55.1	47.8	47.8	48	42.4	37.8	32.2	31.5	28.3	30.7
AM183329_HDV6	69.3	72	69.6	76.3		71.9	72.3	62.4	59.2	52.8	54.5	47.9	47.8	48.1	41.8	35.7	32.1	31.4	27.7	31.2
AM183333_HDV7	68	72.6	70.2	77.3	72.6		75.4	60.9	58.7	51.1	52.1	46.9	46.9	47.3	42.8	36.3	31.2	31	27.9	30.1
AM183327_HDV8	68.9	72.4	75.7	79.9	75.8	75.4		61.5	58.9	52.2	53.6	47.6	46.9	47.5	42.1	35.7	31.1	31.1	26.3	30.4
AB037948_HDV3	66.8	63.2	66.1	65.7	64.6	63.3	64.7		55	50.6	52	46.5	46.1	46.7	41.4	35.7	30.1	29.8	26.7	30.1
OvirDV	61	64.1	69.1	65.1	66.1	62.7	65.6	63		55.8	61.6	47	47.4	47.1	42.7	35.8	31.2	31.1	25.7	28.2
MmonDV	55.4	57.4	60.3	56.4	54.7	54.5	54.9	59.4	67		56.9	45.8	46.2	45.9	43.5	33.7	28.5	29.4	23.4	27.4
DrDV-A	62.6	63.6	64.4	62.6	61.4	60.1	60.5	60.4	74.7	70.1		47	47.4	47.6	43.3	36.3	29	28.5	24.4	27.5
PmacDV	54.3	53.8	55.6	52.3	52.6	52.9	54.3	53.7	57.7	55.1	56.1		95.4	71.2	44.4	33.7	29	30	24.7	26.2
MK598009_PsemDV	54.8	53.8	56.1	52.8	54.2	53.4	54.8	54.7	58.2	55.1	56.1	96.9		72.6	43.8	33.7	29.4	30.1	24.2	26.3
DrDV-B_CAJ1	52.8	51.8	53.1	51.8	51.6	49.9	52.3	54.2	56.1	55.1	56.7	79.6	81.1		43.7	34.5	29.7	29.6	24.7	27.3
NC_040729_SnakeDV	50	47.2	47.7	46.2	47.8	45.9	45.2	47.8	50.8	50.3	51.8	53	54.5	53		34.2	28.7	27.1	23.2	26.2
NC_040845_AvianDV	40.7	36.2	39.4	36.7	35.7	37.9	36.2	38.9	37.2	39.9	38.3	34.7	35.3	37.9	34.9		25.7	30.4	24.9	25.4
MN031240_FishDV	19.8	18.9	20.9	19.9	20.4	18.5	20.4	19.9	23.4	22.5	20.9	19.3	19.3	21.3	17.1	16.1		24.5	22.1	24.4
MN031239_NewtDV	21.8	20.6	19.7	20.6	18.9	19	20.2	20.2	19.6	18.8	20.1	16.9	17.3	17.8	16.7	21.1	14.7		22.9	24.3
MK962760_ToadDV	20.7	24.6	20.9	20.9	20.4	21	19.3	19.9	18.2	20.4	20.9	20.2	20.2	18.6	18.4	20.1	16.8	16.2		19.8
MK962759_TermiteDV	20	18.2	20.1	21.6	19.7	22.2	21.1	21.1	20.1	19.7	21.6	19	21	20.5	18.8	21.5	10.8	14.9	10.4	

Fig. S1. Genetic distances matrices showing representative deltavirus sequences with percent nucleotide identities between genomes (upper triangle) and percent amino acid identities between complete DAg sequences (lower triangle). Darker shading indicates higher percentage identity between two deltaviruses.

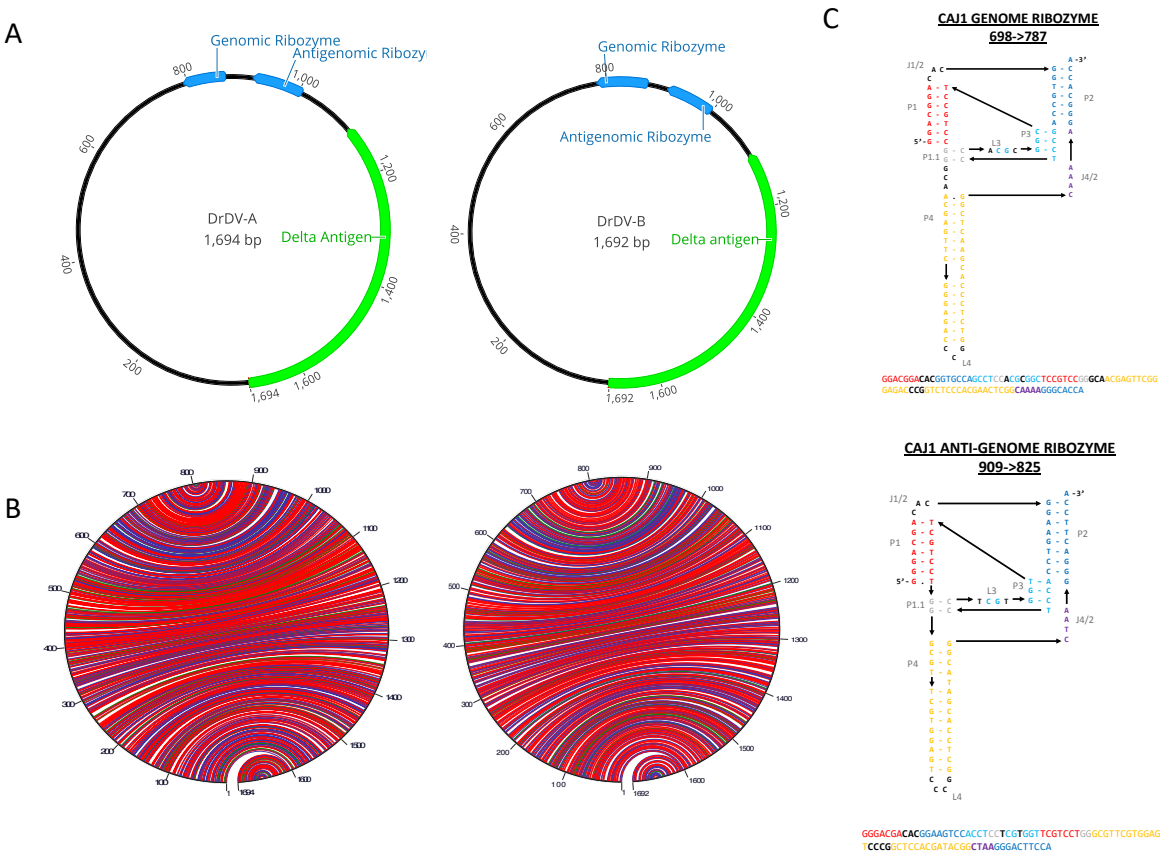

Fig. S2. DrDV genomes exhibit characteristics common to deltaviruses. (**A**) The locations of the delta antigen open reading frame (green) and genomic/antigenomic ribozymes (blue) are shown along the circular genomes of DrDV-A and DrDV-B (CAJ1 shown as an example of DrDV-B). (**B**) Intramolecular base pairing for DrDVs depicted as lines connecting points on the circular genome – G-C pairs are red, A-U pairs are blue, G-U pairs are green, other pairs are yellow. (**C**) Genomic and antigenomic ribozyme secondary structures are shown along with genome location for genome CAJ1. Complementary regions are shown in the same color, and structures are depicted in the style of Webb & Luptak to facilitate comparison with ribozymes from previous studies (3, 5, 6). For further comparison, the ribozyme structures presented in (4) are based on a consensus ribozyme sequence created from an alignment of all deltavirus and deltavirus-like ribozymes. Unlike ribozyme sequences in some other deltavirus genomes, we do not observe a shortening of the J1/2 loop in the DrDV genomic ribozyme compared to the anti-genomic ribozyme, with the sequence CAC present in both.

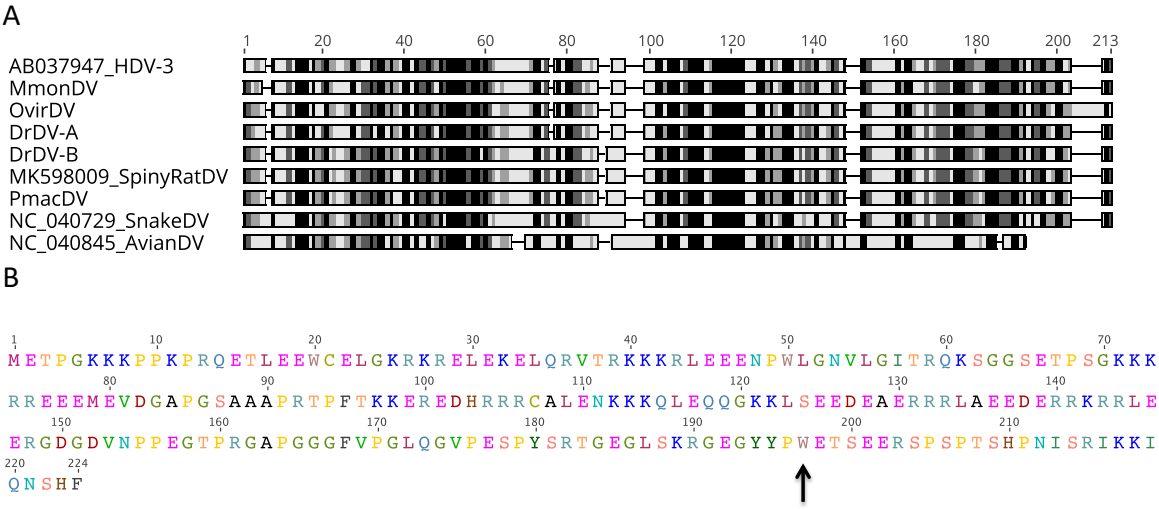

Fig. S3. Characterization of DrDV delta antigen proteins. (**A**) Alignment of delta antigen protein sequences for mammalian, snake and avian deltaviruses. Shading indicates level of similarity across all sequences, with regions of highest identity in black. (**B**) Putative sequence of the large DAg for the DrDV-B virus from the site CAJ1. The RNA editing site is marked with a black arrow; UAG has been edited to UGG yielding a tryptophan residue (W).

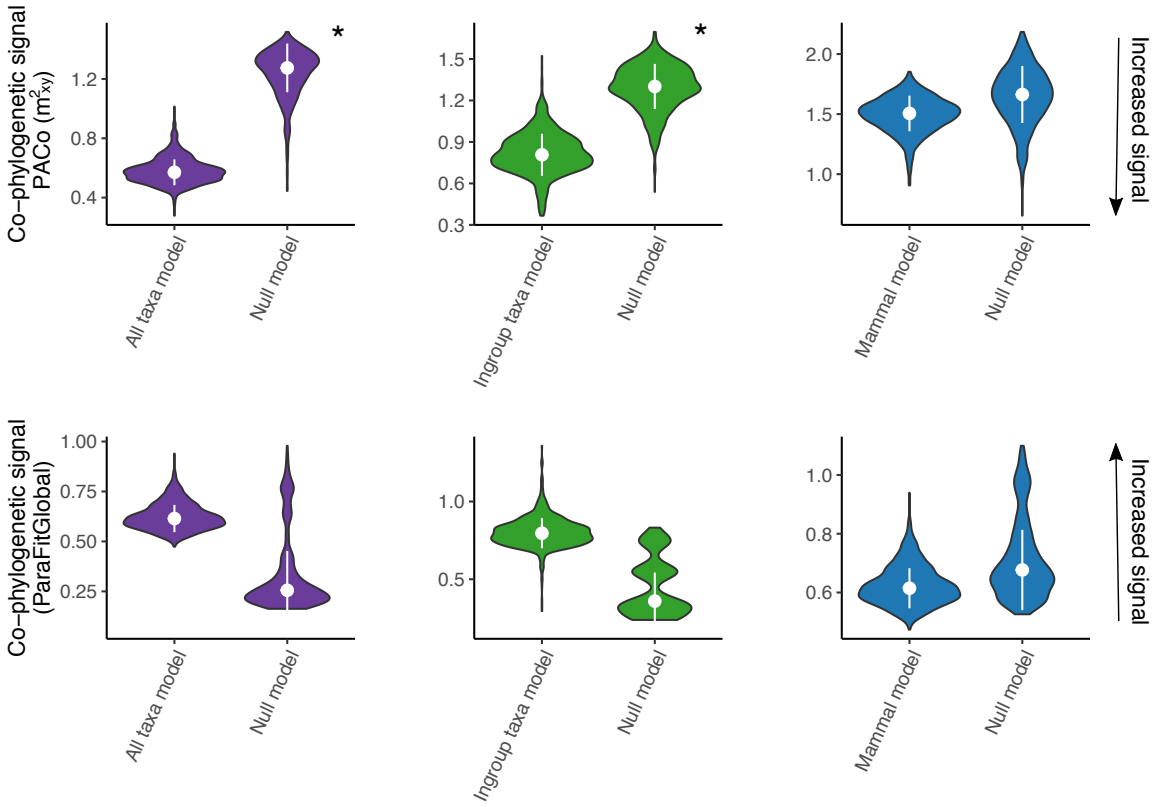

Fig. S4. Co-phylogenetic signal in subsets of the deltavirus phylogeny. Violin plots show the degree of dependence of 1,000 phylogenies from the posterior of the StarBeast analysis (Fig. S5) on the host phylogeny relative to null models, with the median and standard deviation. Data subsets are colored as in Fig. 2B (All taxa: purple+green+blue, Ingroup: green+blue, mammals: blue) Distributions are shown for analyses performed using PACo (top row) and ParaFit (bottom row). Asterisks show significant dependence of the virus phylogeny on the host phylogeny ($P < 0.05$). Note that lower values of the empirical model relative to the null model represent increased signal of co-phylogeny in PACo while higher values represent increased signal of co-phylogeny in ParaFit.

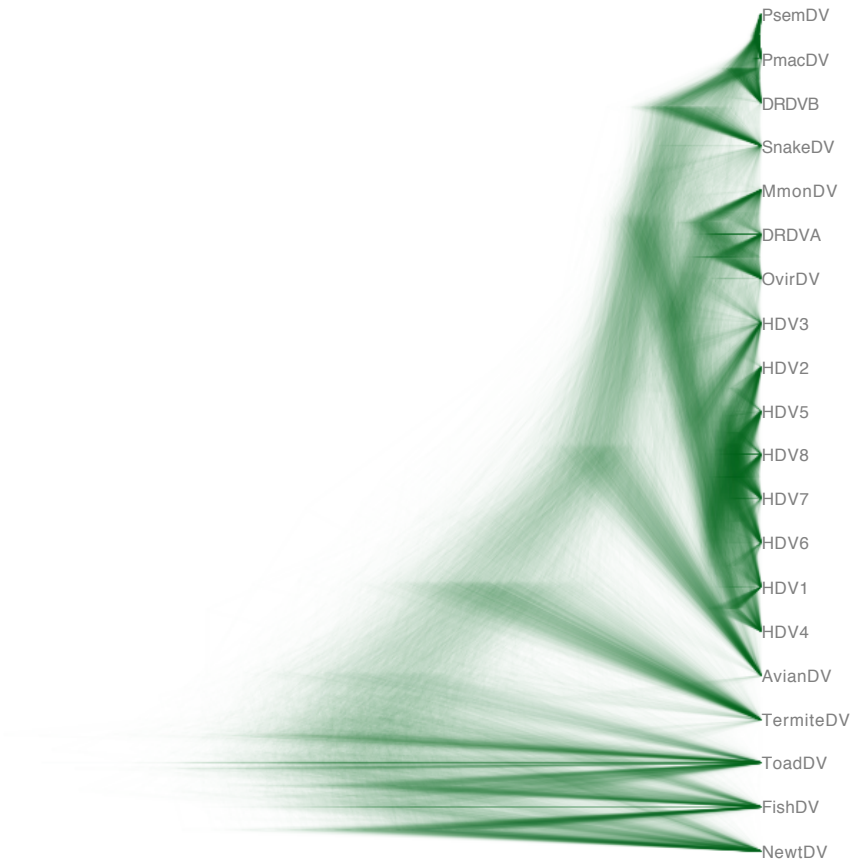

Fig. S5. Uncertainty of deep relationships in the deltavirus phylogeny. The DensiTree shows the distribution of 1,000 posterior trees from the StarBeast analysis, highlighting uncertainties in the evolutionary relationships among divergent deltavirus-like taxa.

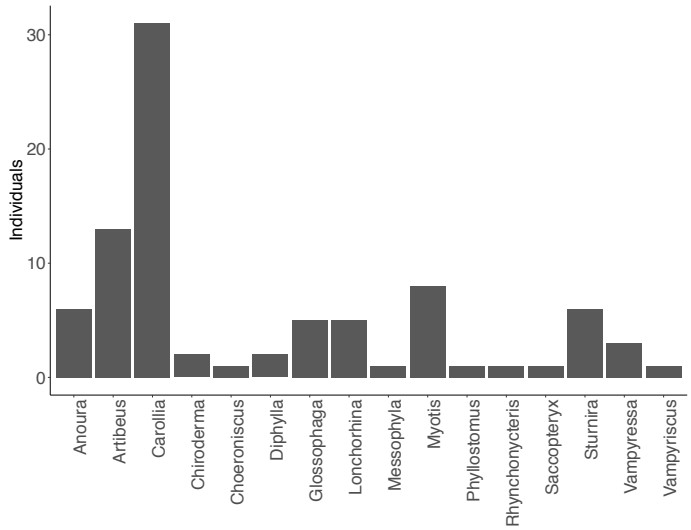

Fig. S6. Counts of non-*D. rotundus* bat species saliva swabs individually screened by RT-PCR for DrDV-B. Bars group bats by genus.

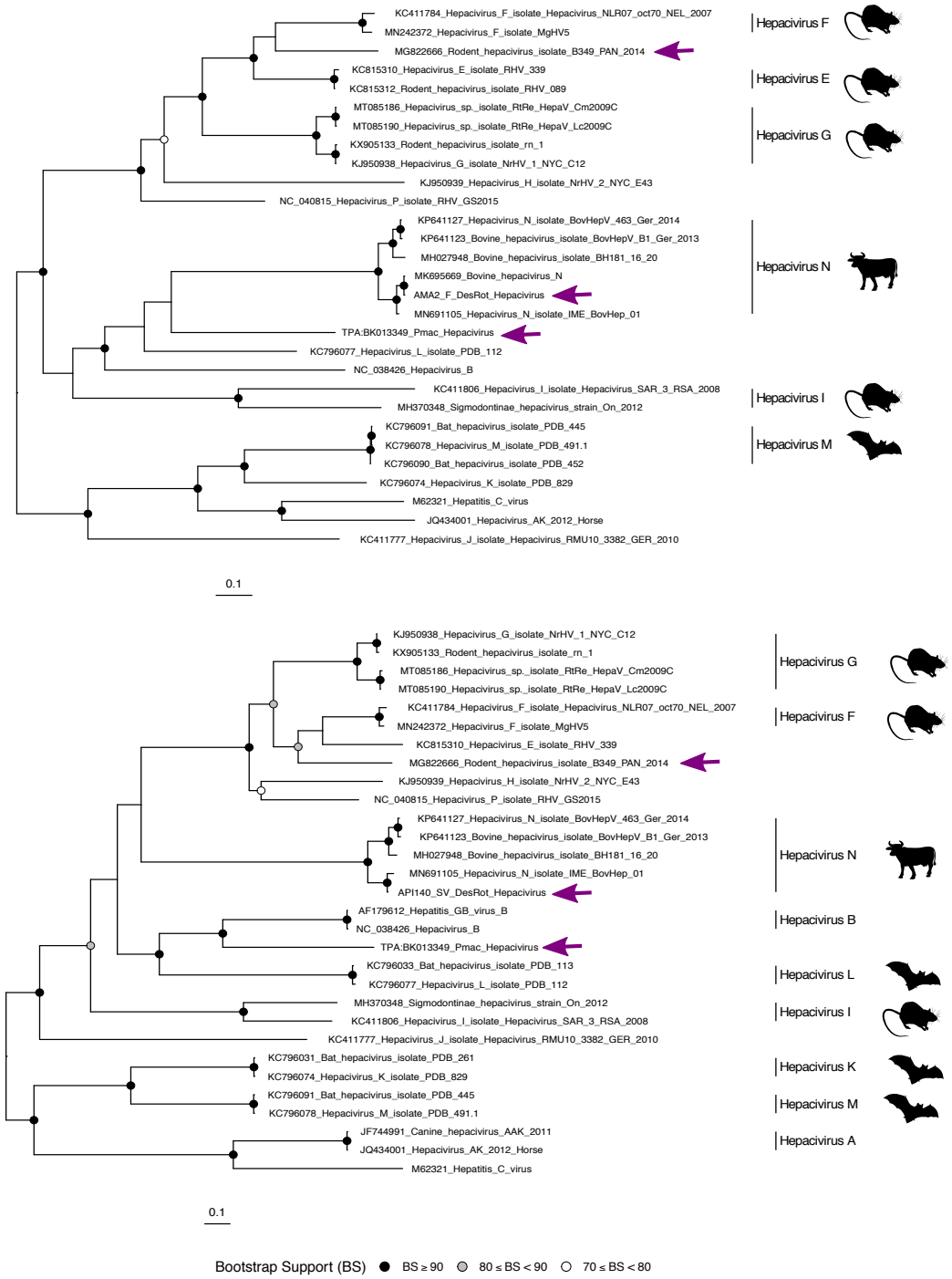

Fig. S7. Relationships between hepaciviruses from deltavirus-positive hosts. Phylogenies shown are based on amino acid alignments of the hepacivirus genes NS3 (upper) and NS5B (lower). Hepacivirus species with multiple representatives are denoted with vertical lines. Silhouettes show host associations for key hepacivirus species. Purple arrows indicate hepaciviruses detected in deltavirus-positive hosts (*Peropteryx macrotis*, *Desmodus rotundus*, and *Proechimys semispinosus*). Maximum likelihood phylogenies with 1,000 bootstrap replicates were generated using IQTree (7) using the best fit models LG+F+I+G4 (NS3) and LG+I+G4 (NS5B) selected by ModelFinder within IQTree 2 (8).

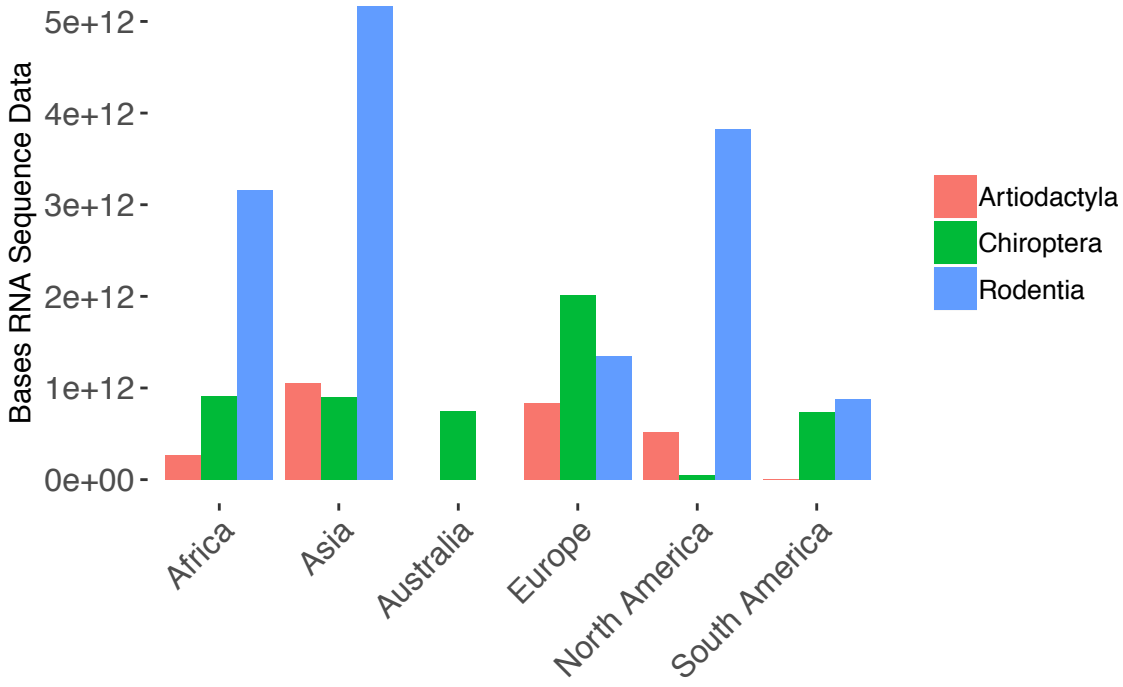

Fig S8. Continent level geographic biases in RNA sequence data examined by Serratus. Bars are colored by mammalian order; data shown are limited to the three orders in which deltaviruses were detected.

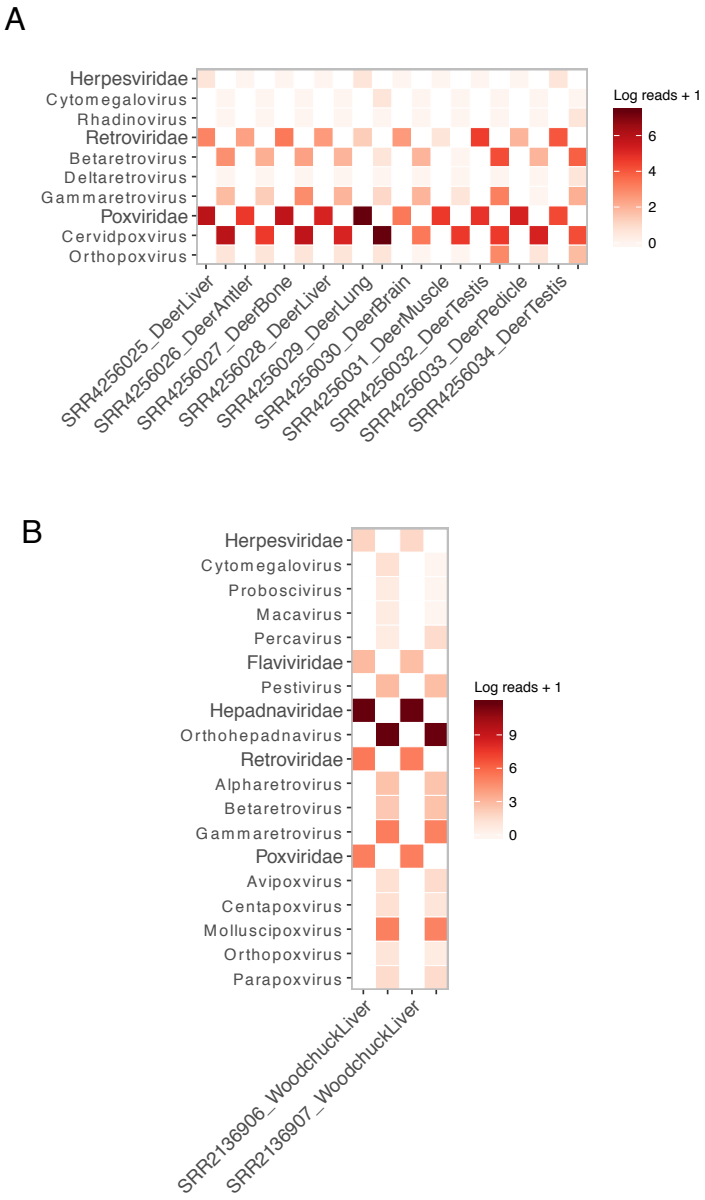

Fig S9. Candidate helper viruses for the OvirDV and MmonDV datasets. Mammal-infecting viral communities are shown for (A) *O. virginianus* libraries sequenced by RNASeq from (9), several of which contained OvirDV and (B) two *M. monax* samples infected with MmonDV from (10). Viral families (in larger font) and genera are shown in adjacent columns for each sample, with families on the left and genera on the right.

Supplementary Tables

Table S1. Pooled bat saliva samples from Peru analyzed by metagenomic sequencing.

Genus	Species	Individuals in pool	Raw reads	Deltavirus contig length (nt)
Carollia	*perspicillata*	10*	28,700,978	N
Glossophaga	*soricina*	5	24,079,752	N
Desmodus	*rotundus*	10	28,946,275	921†
Diphylla	*ecaudata*	2	25,023,095	N
Anoura	*geoffroyi peruana*	6	18,569,505	N
Artibeus	*lituratus obscurus planirostris fraterculus*	10	14,966,399	N
Myotis	*oxyotus unidentified sp*	8	19,934,479	N
Sturnira	*erythromos unidentified sp*	6	11,348,995	N
Vampyressa/ Vampyriscus	*bidens unidentified sp*	4	13,734,389	N
Rare species	*Chiroderma trinitatum Chiroderma salvini Choeroniscus minor Rhynchonycteris naso Saccopteryx bilineata Messophyla macconelli Phyllostomus discolor Rhinophylla pumilio*	8	16,746,795	N

*Pool included individual CP-1 in which DrDV-B was detected by RT-PCR
†Pool was identical to CAJ1 pool where DrDV was initially discovered, confirming the ability to detect deltaviruses when they are known to be present

Table S2. Deltavirus positive cohorts evaluated by mapping reads from related libraries to novel deltavirus genomes.

SRA accession	Pool ID	Host[§]	DV Reads[¶]
ERR2756783	AAC_H_F	*Desmodus rotundus*	2 (DrDV-A)
ERR2756784	AAC_H_SV*	*Desmodus rotundus*	189 (DrDV-A) 18 (DrDV-B)
ERR2756785	AAC_L_F	*Desmodus rotundus*	0
ERR2756786	AAC_L_SV	*Desmodus rotundus*	0
ERR2756787	AMA_L_ F_NR	*Desmodus rotundus*	0
ERR2756788	AMA_L_F_R	*Desmodus rotundus*	0
ERR2756789	AMA_L_SV	*Desmodus rotundus*	0
ERR2756790	CAJ_L_F_NR	*Desmodus rotundus*	0
ERR2756791	CAJ_L_F_R	*Desmodus rotundus*	0
ERR2756792	CAJ_L_SV	*Desmodus rotundus*	4
ERR2756793	CAJ_H_F_1	*Desmodus rotundus*	0
ERR2756794	CAJ_H_F_2	*Desmodus rotundus*	4
ERR2756795	CAJ_H_SV*	*Desmodus rotundus*	169
ERR2756796	HUA_H_F	*Desmodus rotundus*	2
ERR2756797	HUA_H_SV	*Desmodus rotundus*	0
ERR2756798	LMA_L_F_NR	*Desmodus rotundus*	0
ERR2756799	LMA_L_F_R	*Desmodus rotundus*	0
ERR2756800	LMA_L_SV_NR	*Desmodus rotundus*	45
ERR2756801	LMA_L_SV_R*	*Desmodus rotundus*	320
ERR2756802	LR_L_F_NR	*Desmodus rotundus*	0
ERR2756803	LR_L_F_R	*Desmodus rotundus*	0
ERR2756804	LR_L_SV	*Desmodus rotundus*	0
ERR3569452	AMA2_H	*Desmodus rotundus*	0
ERR3569453	AMA2_SV	*Desmodus rotundus*	0
ERR3569454	API1_H	*Desmodus rotundus*	0
ERR3569455	API1_SV	*Desmodus rotundus*	0
ERR3569456	API17_H	*Desmodus rotundus*	0
ERR3569457	API17_SV	*Desmodus rotundus*	2 (DrDV-A)
ERR3569458	API140_H	*Desmodus rotundus*	32
ERR3569459	API140_SV	*Desmodus rotundus*	0
ERR3569460	API141_H	*Desmodus rotundus*	0
ERR3569461	API141_SV	*Desmodus rotundus*	0
ERR3569462	AYA1_H	*Desmodus rotundus*	2
ERR3569463	AYA1_SV	*Desmodus rotundus*	0
ERR3569464	AYA7_H	*Desmodus rotundus*	0
ERR3569465	AYA7_SV	*Desmodus rotundus*	18
ERR3569466	AYA11_H	*Desmodus rotundus*	0

ERR3569467	AYA11_SV*	*Desmodus rotundus*	591
ERR3569468	AYA12_H	*Desmodus rotundus*	0
ERR3569469	AYA12_SV	*Desmodus rotundus*	2
ERR3569470	AYA14_H	*Desmodus rotundus*	0
ERR3569471	AYA14_SV*	*Desmodus rotundus*	228 (DrDV-A) 2 (DrDV-B)
ERR3569472	AYA15_H	*Desmodus rotundus*	0
ERR3569473	AYA15_SV	*Desmodus rotundus*	0
ERR3569474	CAJ1_H	*Desmodus rotundus*	0
ERR3569475	CAJ1_SV*	*Desmodus rotundus*	172
ERR3569476	CAJ2_H	*Desmodus rotundus*	4
ERR3569477	CAJ2_SV†	*Desmodus rotundus*	0
ERR3569478	CAJ4_H	*Desmodus rotundus*	0
ERR3569479	CAJ4_SV†	*Desmodus rotundus*	0
ERR3569480	CUS8_H	*Desmodus rotundus*	0
ERR3569481	CUS8_SV	*Desmodus rotundus*	4
ERR3569482	HUA1_H	*Desmodus rotundus*	0
ERR3569483	HUA1_SV	*Desmodus rotundus*	0
ERR3569484	HUA2_H	*Desmodus rotundus*	5
ERR3569485	HUA2_SV	*Desmodus rotundus*	0
ERR3569486	HUA3_H	*Desmodus rotundus*	0
ERR3569487	HUA3_SV	*Desmodus rotundus*	0
ERR3569488	HUA4_H	*Desmodus rotundus*	0
ERR3569489	HUA4_SV	*Desmodus rotundus*	0
ERR3569490	LMA5_H	*Desmodus rotundus*	0
ERR3569491	LMA5_SV	*Desmodus rotundus*	8
ERR3569492	LMA6_H	*Desmodus rotundus*	0
ERR3569493	LMA6_SV*	*Desmodus rotundus*	173
ERR3569494	LR2_H	*Desmodus rotundus*	0
ERR3569495	LR2_SV	*Desmodus rotundus*	0
ERR3569496	LR3_H	*Desmodus rotundus*	0
ERR3569497	LR3_SV	*Desmodus rotundus*	0
SRR7910142	Pm_03	*Peropteryx macrotis*	0
SRR7910143	Pm_01*	*Peropteryx macrotis*	346
SRR7910144	Pm_02	*Peropteryx macrotis*	2
SRR7910145	Nl_02‡	*Nyctinomops laticaudatus*	0
SRR7910146	Nl_03‡	*Nyctinomops laticaudatus*	0
SRR7910147	Mk_01‡	*Myotis keaysi*	0
SRR7910148	Mk_02‡	*Myotis keaysi*	0
SRR7910149	Mm_02‡	*Mormoops megalophylla*	0
SRR7910150	Mm_03‡	*Mormoops megalophylla*	0

SRR7910151	Aj_03‡	*Artibeus jamaicensis*	0
SRR7910152	Mm_01‡	*Mormoops megalophylla*	0
SRR7910153	Aj_01‡	*Artibeus jamaicensis*	0
SRR7910154	Aj_02‡	*Artibeus jamaicensis*	0
SRR7910155	Mk_03‡	*Myotis keaysi*	0
SRR7910156	Nl_01‡	*Nyctinomops laticaudatus*	0
SRR4256025	Deer-Liver-1-male	*Odocoileus virginianus*	0
SRR4256026	Deer-Antler-2-male	*Odocoileus virginianus*	12
SRR4256027	Deer-Bone-1-male	*Odocoileus virginianus*	0
SRR4256028	Deer-Liver-2-male	*Odocoileus virginianus*	0
SRR4256029	Deer-Lung-1-male	*Odocoileus virginianus*	0
SRR4256030	Deer-Brain-2-male	*Odocoileus virginianus*	12
SRR4256031	Deer-Muscle-2-male	*Odocoileus virginianus*	14
SRR4256032	Deer-Testis-1-male	*Odocoileus virginianus*	0
SRR4256033	Deer-Pedicle-male*	*Odocoileus virginianus*	9265
SRR4256034	Deer-Testis-2-male	*Odocoileus virginianus*	7

* Pools in which full deltavirus genomes were detected
† Pools in which DrDV was detected in the saliva of one or more individuals in the pool by RT-PCR, but were negative for deltavirus detection through metagenomics
‡Species in which deltaviruses were not detected but which came from the same study
§*D. rotundus* samples represent saliva (SV) and fecal (F/H) samples pooled across multiple individuals from different sites. Samples from other Neotropical bats (*P. macrotis, N. laticaudatus, M. keaysi, M. megalophylla, A. jamaicensis*) represent liver samples from unique individuals. *O. virginianus* samples represent different tissues pooled across multiple individuals. Read mapping of samples from different individuals and time points to the MmonDV genome is described in (4)
¶Samples from *D. rotundus* were mapped to DrDV-A and DrDV-B genomes. All *D. rotundus* read counts refer to DrDV-B genomes unless specifically noted as DrDV-A. In the case of libraries with matches to both, the number of reads mapping is broken down by DrDV-genome. Samples from *P. macrotis, N. laticaudatus, M. keaysi, M. megalophylla, A. jamaicensis* were mapped to the PmacDV genome. Samples from *O. virginianus* were mapped to OvirDV.

Table S3. Summary statistics for bat deltavirus genomes and protein domain homology analysis of predicted DrDV small delta antigens from saliva metagenomic pools.

Site	Lineage	Genome (nt)	GC content (%)	Intramolecular base pairing (%)	Delta antigen (aa)	Hhpred top hit	Probability top hit	e-value	Identity top hit (%)	Genbank accession
AYA14	DrDV-A	1694	55	73.8	194	Oligomerization domain of hepatitis delta antigen	99.86	2.8e-25	59	MT649207
AYA11	DrDV-B	1692	54.3	75.3	196		99.86	5.60E-25	45	MT649206
CAJ1	DrDV-B	1692	53.8	74.3	196		99.86	5.40E-25	45	MT649208
LMA6	DrDV-B	1694	54.3	74.6	196		99.85	8.30E-25	45	MT649209

Table S4. Primers used to screen samples for DrDV by RT-PCR and HBV by PCR.

Primer	PCR Round	Sequence (5´-3´)
DrDV-A		
DrDV_F1_GenoA	1&2	AGGGGTCTTTTTGGGAAATT
DrDV_R1_GenoA	1	AAGAAGAAGCAACTATCCGG
DrDV_R2_GenoA	2	CATCCAAGAGACCAAGAGAG
DrDV-B		
DrDV_F1_GenoB	1	TTCCCTTGYTGCTCCAGTTG
DrDV_R1_GenoB	1	CGGTAAGAAGAAACCTCCAA
DrDV_F2_GenoB	2	CCAGTTGTTTCTTCTTGTTCTC
DrDV_R2_GenoB	2	AAAAAGAAAGAGAGAACTGGAAAAA
DrDV Delta Antigen		
DeltaAntigenF1_GenoB	1	TCTGGTCTTATCTTTCTTACCTTAT
DeltaAntigenR1_GenoB	1	AAACCTTCCTTTATTCTATTTCGAA
DeltaAntigenR1_GenoA	1	CCTTTACCTTTAATTCTCTTGGTAA
DeltaAntigenF1_GenoA	1	GCCTCGAATAATAAGAAGAAAATTT
HBV Primers*		
HBV-F248	1&2	CTAGATTBGTGGTGGACTTCTCTCA
HBV-R397	2	GATARAACGCCGCAGATACATCCA
HBV-R450a	1	TCCAGGAGAACCAAYAAGAAAGTGA
HBV-R450b	1	TCCAGGAGAACCAAYAAGAAGATGA

*Primer sequences and PCR protocol described in (11)

Table S5. Colony level demographic characteristics and PCR-based screening results of vampire bat blood and saliva for DrDV and HBV.

			DrDV-A		DrDV-B		HBV	
Colony	Prop Male*	Prop Adult†	Saliva	Blood	Saliva	Blood	Saliva	Blood
AYA1	0.6	1	0/20	0/20	3/20	0	0/3	0/20
AYA11	0.6	0.95	0/20	0/20	2/20	0	0/3	0/20
AYA14	0.4	0.65	1/20	0/20	4/20	0	0/8	0/20
AYA15	0.55	0.75	0/20	0	0/20	0	0	0
CAJ1	0.75	0.9	0/20	0	5/20	0/20	0/10	0/20
CAJ2	0.55	0.95	0/20	0	6/20	6/20	0/6	0/20
CAJ3	0.7	1	0/20	0	4/20	0	0/4	0
CAJ4	0.35	0.75	0/20	0	2/20	0	0/2	0
LMA4	0.65	0.75	0/20	0	0/20	0	0/1	0
LMA5	0.65	0.9	0/20	0	5/20	0	0/5	0
LMA6	0.35	1	0/20	0	7/20	4/20	0/9	0/19
LMA12	0.5	0.9	0/20	0	3/20	0	0/3	0
Total	-	-	1/240	0/60	41/240	10/60	0/54	0/119

*Proportion of males at each colony (alternative is females)
†Proportion of adults at each colony (alternatives are juveniles or subadults)

Table S6. Test of association between DrDV-B phylogeny and sample location at the regional (department) and colony level.

Level	Index	Observed value (95% CI)	Null value (95% CI)	p-value
Region	AI*	0.22 (0-0.58)	2.46 (1.95-2.94)	0
	PS†	4 (3-5)	17.73 (15.41-19.35)	0
	MC‡ (LMA)	9.29 (5-14)	1.97 (1.41-2.98)	0.001
	MC‡ (CAJ)	11.24 (9-19)	2.81 (2.12-3.94)	0.001
	MC‡ (AYA)	2.67 (1-5)	1.26 (1-1.96)	0.02
Colony	AI*	2.23 (1.69-2.76)	3.63 (3.21-3.92)	0
	PS†	19.19 (18-20)	29.23 (27.45-30.81)	0
	MC‡ (LMA6)	4.98 (5-5)	1.18 (1-1.94)	0.001
	MC‡ (LMA5)	2.52 (1-3)	1.13 (1-1.43)	0.002
	MC‡ (CAJ2)	3.21 (2-6)	1.57 (1.14-2.39)	0.01
	MC‡ (CAJ1)	1 (1-1)	1.13 (1-1.53)	1
	MC‡ (AYA1)	1.09 (1-2)	1.01 (1-1.05)	1
	MC‡ (CAJ3)	1.01 (1-1)	1.07 (1-1.32)	1
	MC‡ (AYA11)	1.03 (1-1)	1.01 (1-1.05)	1
	MC‡ (AYA14)	1.05 (1-2)	1.04 (1-1.15)	1
	MC‡ (CAJ4)	1.16 (1-2)	1.01 (1-1.05)	1
	MC‡ (LMA12)	1 (1-1)	1.04 (1-1.15)	1

*Association Index
† Parsimony Score
‡ Monophyletic Clade size

Dataset S1 (separate file). Merged Serratus and Pantheria datasets. Data S1 is an excel file containing merged datasets used to evaluate geographic and taxonomic biases in SRA searches.

Dataset S2 (separate file). Full delta antigen alignment. Data S2 is a fasta file which is a trimmed alignment of the amino acid sequence of the full delta antigen used to evaluate relationships between deltavirus representatives from different hosts.

Dataset S3 (separate file). Vampire bat deltavirus infections. Data S3 is an excel file containing individual level infection data for vampire bats screened by PCR for DrDV-B along with demographic data used in statistical analyses.

Dataset S4 (separate file). DrDV-B partial delta antigen alignment. Data S4 is a fasta file which is a trimmed alignment of the nucleotide sequence of the partial delta antigen fragment used to evaluate relationships between deltaviruses detected in vampire bats from different regions of Peru. A DrDV-B sequence detected in a co-roosting *C. perspicillata* individual is also included in this alignment.

Supplementary References

1. Casey JL, Bergmann KF, Brown TL, Gerin JL (1992) Structural requirements for RNA editing in hepatitis delta virus: evidence for a uridine-to-cytidine editing mechanism. *Proc Natl Acad Sci USA* 89(15):7149–7153.

2. Hwang SB, Lai M (1993) Isoprenylation Mediates Direct Protein-Protein Interactions Between Hepatitis Large Delta-Antigen and Hepatitis-B Virus Surface-Antigen. *J Virol* 67(12):7659–7662.

3. Wille M, et al. (2018) A Divergent Hepatitis D-Like Agent in Birds. *Viruses* 10(12):720–9.

4. Edgar RC, et al. (2020) Petabase-scale sequence alignment catalyses viral discovery. *bioRxiv*:1–30.

5. Webb C-HT, Lupták A (2014) HDV-like self-cleaving ribozymes. *RNA Biol* 8(5):719–727.

6. Hetzel U, et al. (2019) Identification of a Novel Deltavirus in Boa Constrictors. *mBio* 10(2):1447–8.

7. Minh BQ, et al. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37(5):1530–1534.

8. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler von A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Meth* 14(6):587–589.

9. Seabury CM, et al. (2011) Genome-Wide Polymorphism and Comparative Analyses in the White-Tailed Deer (Odocoileus virginianus): A Model for Conservation Genomics. *PLOS ONE* 6(1):e15811–9.

10. Fletcher SP, et al. (2015) Intrahepatic Transcriptional Signature Associated with Response to Interferon-α Treatment in the Woodchuck Model of Chronic Hepatitis B. *PLoS Pathog* 11(9):e1005103–30.

11. 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 USA* 110(40):16151–16156.