# Identification of novel avian and mammalian deltaviruses provides new insights into deltavirus evolution

Masashi Iwamoto[1,2], Yukino Shibata[3], Junna Kawasaki[4,5], Shohei Kojima[4,6], Yung-Tsung Li[7], Shingo Iwami[2], Masamichi Muramatsu[1], Hui-Lin Wu[7,8], Kazuhiro Wada[3], Keizo Tomonaga[4,5,9], Koichi Watashi[1,10], Masayuki Horie[4,11,*]

[1]*Department of Virology II, National Institute of Infectious Diseases, Tokyo, Japan*

[2]*Mathematical Biology Laboratory, Department of Biology, Faculty of Sciences, Kyushu University, Fukuoka, Japan*

[3]*Graduate School of Life Science, Hokkaido University, Sapporo, Hokkaido, Japan*

[4]*Institute for Frontier Life and Medical Science, Kyoto University, Kyoto, Japan*

[5]*Department of Mammalian Regulatory Network, Graduate School of Biostudies, Kyoto University, Kyoto, Japan*

[6]*Genome Immunobiology RIKEN Hakubi Research Team, RIKEN Cluster for Pioneering Research*

[7]*Hepatitis Research Center, National Taiwan University Hospital, Taipei, Taiwan*

[8]*Graduate Institute of Clinical Medicine, National Taiwan University College of Medicine*

[9]*Department of Molecular Virology, Graduate School of Medicine, Kyoto University, Kyoto, Japan*

[10]*Department of Applied Biological Sciences, Tokyo University of Science, Noda, Japan*

[11]*Hakubi Center for Advanced Research, Kyoto University*

*Corresponding author

Masayuki Horie, DVM, PhD

Hakubi Center for Advanced Research, Kyoto University

53 Kawahara-cho, Shogo-in, Sakyo, Kyoto 606-8507, Japan

26 Email: horie.masayuki.3m@kyoto-u.ac.jp

27

**Abstract**

Hepatitis delta virus (HDV) is a satellite virus that requires hepadnavirus envelope proteins for its transmission. Although recent studies identified HDV-related deltaviruses in certain animals, the evolution of deltaviruses, such as the origin of HDV and the mechanism of its coevolution with its helper viruses, is unknown, mainly because of the phylogenetic gaps among deltaviruses. Here we identified novel deltaviruses of passerine birds, woodchucks, and white-tailed deer by extensive database searches and molecular surveillance. Phylogenetic and molecular epidemiological analyses suggest that HDV originated from mammalian deltaviruses and the past interspecies transmission of mammalian and passerine deltaviruses. Further, metaviromic and experimental analyses suggest that the satellite-helper relationship between HDV and hepadnavirus was established after the divergence of the HDV lineage from non-HDV mammalian deltaviruses. Our findings enhance our understanding of deltavirus evolution, diversity, and transmission, indicating the importance of further surveillance for deltaviruses.

## Introduction

Hepatitis delta virus (HDV) is the only member of the genus *Deltavirus* which is not assigned to a family (1). The HDV genome is an approximately 1.7-kb circular, negative single-stranded RNA, harboring a single open reading frame (ORF) encoding the small and large hepatitis delta antigens (S-HDAg, 24 kDa and L-HDAg, 27 kDa) that are translated from the same transcriptional unit via RNA-editing of the stop codon, which is catalyzed by the host protein ADAR1 (2-5). The 19 amino acid residue extension of the C-terminal region of L-HDAg contains a farnesylation site required to interact with helper virus envelope proteins (6). The genome structure of HDV is unique in that it has genomic and antigenomic ribozymes, which are essential for its replication (7, 8), and is highly self-complementarity, generating a rod-like structure (9-11). Although HDV can autonomously replicate, it requires an envelope protein of other "helper" viruses to produce infectious virions. Hepatitis B virus (HBV) (family *Hepadnaviridae*) provides the envelope proteins required for HDV transmission between humans (12). Approximately 15–20 million people worldwide are estimated to be infected with HDV among 350 million HBV carriers (13). Compared with mono-infection with HBV, coinfection of HDV and HBV accelerates the pathogenic effects of HBV, such as severe or fulminant hepatitis and progression to hepatocellular carcinoma, through unknown mechanisms (14).

The evolutionary origin of HDV presents an enigma. However, recent discoveries of deltaviruses of vertebrate and invertebrate species (15-18), significantly changed our understanding of deltavirus evolution. These non-HDV deltaviruses are distantly related to HDV but may share the same origin because of their similarly structured circular RNA genomes (approximately 1.7 kb), which encode DAg-like proteins, possess ribozymes

66 sequences, and are highly self-complementary (15-18). These findings provide clues to
67 the mechanism of deltavirus evolution. For example, a recent study hypothesizes that
68 mammalian deltaviruses codiverged with their host mammalian species (18). However,
69 the few known deltaviruses are highly divergent (15-18). Therefore, the phylogenetic
70 gaps between the deltaviruses must be filled through the identification of putative novel
71 deltaviruses.

72 The discoveries of non-HDV deltaviruses provides insights into the relationships
73 between deltaviruses and their helper viruses. Recently identified non-HDV deltaviruses
74 likely do not coinfect with hepadnaviruses, suggesting the presence of other helper
75 viruses (15-18). This hypothesis is supported by absence of a large isoform of DAg,
76 which is required for the interaction of HDV with the HBV envelope proteins, in rodent
77 deltavirus (18). Further, viral envelope proteins of reptarenavirus and hartmanivirus, but
78 not HBV, confer infectivity upon the snake deltavirus (19). These findings suggest that
79 hepadnaviruses do not serve as helper viruses for non-HDV deltaviruses and that the
80 deltavirus-hepadnavirus relationship is specific to the HDV lineage. However, the large
81 phylogenetic gap between HDV and the few other deltaviruses makes it difficult to assess
82 the hypothesis, raising the importance of further research.

83 In this study, to understand the evolution of deltaviruses, we analyzed publicly
84 available transcriptome data and found novel mammalian and avian deltaviruses. Our
85 phylogenetic analysis suggests that HDVs originated from non-HDV mammalian
86 deltaviruses and does not support the codiversification hypothesis of deltavirus and
87 mammalian evolution. Moreover, *in silico* and experimental analyses, together with
88 previous findings, suggest that the satellite-helper relationship between HDV and
89 hepadnavirus was established after the divergence of the novel non-HDV mammalian

deltaviruses and the HDV lineage. Further, we present evidence for recent interfamily transmission of deltaviruses among passerine birds. Our findings therefore provide novel insights into the evolution of deltaviruses.

## Results

### *Identification of deltavirus-related sequences in avian and mammalian transcriptomes*

We first assembled 46,359 RNA-seq data of birds and mammals. Using the resultant contigs as queries, we identified five deltavirus-related contigs in the SRA data of birds and mammals, including the zebra finch (*Taeniopygia guttata*), common canary (*Serinus canaria*), Gouldian finch (*Erythrura gouldiae*), Eastern woodchuck (*Marmota monax*), and white-tailed deer (*Odocoileus virginianus*). We named the deltavirus-like sequences Taeniopygia guttata deltavirus (tgDeV), Serinus canaria-associated deltavirus (scDeV), Erythrura gouldiae deltavirus (egDeV), Marmota monax deltavirus (mmDeV), and Odocoileus virginianus (ovDeV), respectively (Table 1).

The amino acid sequences tlanslated from the contigs are 36.0%–66.7% identical to those of the DAg proteins of known deltaviruses (Table 1, Supp Tables 1 and 2). The tgDeV, mmDeV, and ovDeV contigs, which comprise approximately 1,700 nucleotides, encode one ORF with a sequence similar to those of DAg genes of known deltaviruses (Fig 1a, Table 1, Supp Tables 1 and 2). In contrast, the contigs scDeV and egDeV are 761 and 596 nucleotides in length, respectively (Fig. 1b, Supp Tables 1 and 2). Note that the nucleotide sequences of tgDeV and egDeV are 97.7% identical, and we therefore analyzed tgDeV instead of tgDeV and egDeV.

### *Genome structures of novel avian and mammalian deltaviruses*

The three contigs (tgDeV, mmDeV, and ovDeV) are almost identical in length to the full-length genomes of known deltaviruses. We therefore checked for potential circularity of the contigs. Dot-plot analyses revealed that each of both ends of these three contigs is identical (Supp Fig. 1), suggesting that the contigs were derived from circular RNAs. We

118 further mapped the original RNA-seq data to the corresponding circularized contigs using
119 the Geneious mapper, revealing that some of the reads properly spanned the junctions
120 (data not shown), indicating that these contigs are derived from circular RNAs. Therefore,
121 we designated the resultant circular contigs of tgDeV, mmDeV, and ovDeV as full-length
122 novel deltavirus genomes (1,706, 1,712, and 1,690 nucleotides, respectively) (Fig. 1a).
123 These novel genomes are characterized by high self-complementarity, genomic and
124 antigenomic ribozymes, and poly(A) signals, which are conserved among known
125 deltaviral genomes (Fig. 1 and Supp Table 2) (15-18). Further, the predicted secondary
126 structures of the ribozymes are highly similar to those of HDV as well as those of other
127 deltaviruses (Supp Fig. 2).
128
129 ***Characterization of DAg proteins encoded by the novel deltaviruses***
130 We next characterized the putative DAg proteins encoded by the novel deltaviruses. Most
131 of their biochemical features, biologically relevant amino acid residues, and functional
132 domains (15-18) are conserved among the DAg proteins (Fig. 2a). The isoelectric points
133 of DAg proteins from the novel deltaviruses range from 10.35 to 10.63 (Supp Table 2),
134 which are nearly identical to those of known deltaviruses. All the post-translational
135 modification sites in HDAg are conserved among those of all DAg proteins of the novel
136 deltaviruses, except for the serine phosphorylation site on scDeV-DAg (Fig. 2a). The
137 NLS is conserved among the DAg proteins, although location of the predicted NLS of
138 scDeV DAg protein differs (Fig. 2a).
139 We next investigated whether the novel deltaviruses utilize A-to-I RNA-editing. To
140 answer this question, we mapped short reads of the SRA data, which we initially used to
141 detect the deltaviruses, to identify the nucleotide variations among the stop codons. We

found a potential RNA-editing site within the stop codon of the ovDeV-DAg gene, in which there was 0.4% nucleotide variation (5 of 1160 reads) at the second nucleotide position of the stop codon (UAG), all of which were G instead of the consensus nucleotide A (Fig. 2b). The quality scores of the five G variants ranged from 35 to 41 (Supp Fig. 3), which likely exclude the possibility of a sequencing error. This variation may be explained by A-to-I editing by ADAR1, as known for HDV (2). However, possible RNA-editing generates a protein two amino acid residues longer because of a stop codon immediately downstream (Fig. 2c). Further, the C-terminal farnesylation motif (CXXQ) required for the interaction with hepadnaviral envelope proteins (6) was absent from the longer product. These observations suggest that even if RNA-editing occurs, the resultant gene product does not contribute to the interaction with hepadnaviral envelope proteins. Further, we were unable to identify nucleotide variations of the mapped reads at the stop codons in the genomes of tgDeV and mmDeV (data not shown).

***The novel deltaviruses potentially replicate in their hosts***

To determine whether the novel deltaviruses potentially replicate in their respective, putative host species, we evaluated the mapping pattern of viral reads described above. We found that the read depths of the predicted transcribed regions (the DAg coding regions to poly-A signals) were much greater than those of the other genomic regions (Fig. 3), indicating that most viral reads were derived from viral mRNAs. These findings suggest that the novel deltaviruses replicate in their hosts.

The mapping pattern on tgDeV differed slightly from the others. Specifically, although the read depth of the DAg ORF region was higher, the reads represented only 80% of the ORF (Fig. 3a). This trend was apparent in another tgDeV-positive RNA-seq

data (Supp Fig. 4). However, it is not clear whether this is attributed to an artifact or actually reflects the transcription pattern of tgDeV.

***Transmission of tgDeV- and tgDeV-like viruses among passerine birds***

We next investigated whether the novel deltaviruses are transmitted among animal populations. We first analyzed tgDeV infections in birds using RNA-seq data (Table 2 and Supp Table 3). Among 6453 SRA data, tgDeV-derived reads were identified in 34 SRAs, including the SRAs in which tgDeV and egDeV were initially detected. The 34 tgDeV-positive SRA data were obtained from tissues such as blood, kidney, and muscles, suggesting broad tropism and viremia, or systemic infection, or both, with tgDeV.

Further, tgDeV sequences were detected in several bird species such as the black-headed bunting (*Emberiza melanocephala*) and yellow-bellied tit (*Pardaliparus venustulus*). All tgDeV-positive bird species belong to the order Passeriformes. These tgDeV-positive SRA data are included in the nine BioProjects deposited by independent researchers, and thus the birds were likely from different sources. Further, the tgDeV-positive sample in SRR9899549 (BioSample accession: SAMN12493457) is derived from a black-headed bunting caught in the wild. These data suggest that tgDeV (or tgDeV-like viruses) circulate among diverse passerine birds, even in the wild.

During the above analysis, we found that SRA data from the yellow-bellied tit (SRR7244693 and SRR7244695–SRR7244698) contain many reads mapped to the tgDeV genome. Therefore, we extracted the mapped reads of SRA data and performed *de novo* assembly. We obtained a 1707-nt circular complete genome sequence, which we designated pvDeV. The pvDeV nucleotide sequence is 98.2% identical to that of tgDeV, and the properties of its DAg protein sequence are similar to those of tgDeV DAg (Supp

Fig. 5).

We next employed real-time RT-PCR to further evaluate potential deltavirus infections of passerine birds. We analyzed 30 and 5 whole-blood samples from zebra and Bengalese finches (*Lonchura striata* var. *domestica*), respectively, and found that one Bengalese finch was positive for real-time RT-PCR test. To exclude the possibility of contamination of a plasmid used as a control for real-time PCR, we performed RT-PCR using a primer set that distinguishes viral from plasmid amplicons (Figs. 4a and b). We obtained a band of the expected size only from the cDNA sample (Fig. 4c), revealing that the bird was truly positive for a tgDeV-like virus. Therefore, we named this virus lsDeV, and further analysis revealed that its full-length genome nucleotide sequence (1708 nt) is 98.2% and 98.4% identical to those of tgDeV and pvDeV, respectively. Moreover, its genome features are almost identical to those of tgDeV and pvDeV (Supp Fig. 5).

***Evidence for recent interfamily transmission of deltaviruses among passerine birds***

We found that the sequence similarities among the passerine deltaviruses (tgDeV, pvDeV, and ls DeV) (Fig. 4d) were not consistent with evolutionary codivergence. According to the TimeTree (20), deltavirus-positive passerine birds diverged approximately 44 million years ago (Fig. 4e and Supp Fig. 6). Considering the rapid evolutionary rates of HDVs (approximately $10^{-3}$ substitutions per site per year) (21-23), it is unlikely that these viruses codiverged with their hosts. Most likely, interfamily transmission occurred relatively recently among passerine birds.

***Transmission of mmDeV among woodchucks***

We similarly analyzed mmDeV infections using SRA data for the order Rodentia, other

than mice (*Mus musculus*) and rats (*Rattus norvegicus*). Our analysis of 4776 SRA datasets detected mmDeV reads in 20 SRA data of seven woodchucks (Table 2 and Supp Table 3). Although these mmDeV-positive SRA data were contributed by the same research group, the animals were apparently obtained at different times (24, 25), suggesting that mmDeV was transmitted among woodchucks. The mmDeV-positive SRA data are derived from samples of liver or peripheral blood mononuclear cells.

We next used real-time RT-PCR to analyze 81 woodchuck samples (liver, n= 43; serum, n = 38). However, mmDeV was undetectable (data not shown), which may be explained by the absence of mmDeV infection or clearance, low level of infection, or both.

***No evidence of transmission of other deltaviruses***

We next focused on ovDeV and scDeV sequences of ruminant animals and passerine birds, respectively (Table 2 and Supp Table 3). We detected ovDeV-derived reads only in five SRA data. The SRA data were obtained from brain, muscle, testis, pedicles, and antlers, suggesting systemic infection, viremia, or both. However, we were unable to determine if these samples were derived from multiple individuals. We detected scDeV-derived reads only from the SRA data in which the virus was initially detected. We therefore were unable to provide evidence for the transmission of ovDeV and scDeV in their host animals.

***Phylogenetic relationships among deltaviruses***

To decipher the evolutionary relationships among deltaviruses, we conducted a phylogenetic analysis using known and the novel deltavirus sequences discovered here.

We did not include sequences of recently identified fish, toad, newt, termite, and duck-associated deltaviruses because they share very low amino acid sequence identities with the novel deltaviruses as well as with HDVs (Fig. 5a), which may reduce the accuracy of tree (16). We further excluded scDeV for this reason, and we did not include tgDeV-like viruses, because their sequences are nearly identical to that of tgDeV. The reconstructed tree shows that the newly identified tgDeV forms a strongly supported cluster with snake DeV and rodent DeV, although they are distantly related to each other (Fig. 5b). Note that mmDeV and ovDeV are more closely related to HDVs than the other deltaviruses.

***Candidate helper viruses***

To gain insights into helper viruses of the novel deltaviruses, we first analyzed the coexisting viruses in the SRA data using BLASTx. Note that we omitted experimental woodchuck hepatitis virus (WHV) infections associated with the mmDeV-positive woodchuck-derived SRA data (SRR2136864 to SRR2136999). We also excluded viruses that infect invertebrates and endogenous retroviruses as well. These analyses reveal that polyomavirus, bornavirus, and circovirus sequences are present in the deltavirus-positive SRA data of passerine birds (Table 3 and Supp Table 4). Further, we detected contigs with 98%–100% identities to human viruses (human mastadenovirus or mammalian rubulavirus 5) (Supp Table 4), although these may represent contamination, index hopping, or both. Among these viruses, only the genome of bornavirus encodes an envelope protein. Note that scDeV and bornavirus-positive SRA data was obtained from pooled samples (SAMN04260514), and we therefore were unable to determine whether scDeV and canary bornavirus 3 infected the same individual.

We next cross-referenced the mmDeV reads and the metadata, which also provide insights into the mmDeV helper virus. Among 20 mmDeV-positive SRA data, 18 were obtained from animals experimentally infected with the hepadnavirus WHV, which was experimentally shown to serve as a helper virus for HDV (26, 27). However, the other two SRA data (SRR437934 and SRR437938) were derived from animals negative for antibodies against WHV as well as WHV DNA (24). These observations suggest that mmDeV was transmitted to the two animals without WHV and that WHV therefore was not the helper virus for mmDeV that infected these two individuals.

***Replication of tgDeV and mmDeV in human and woodchuck cell lines***

To investigate whether the novel deltavirus sequences are replicable or not, we performed transfection-based assays in cell culture systems. We constructed plasmid expression vectors harboring the minus-strand genome of the tgDeV or mmDeV dimer sequence under the transcriptional control of the CMV promoter (see Materials and Methods). We first determined if the replication initiated by transfecting these plasmids. These plasmids express the minus-strand genome and therefore DAg protein is expressed if the viral genome replicates (19). We transfected the plasmid vectors into Huh7 human hepatic cells and WCH-17 woodchuck hepatic cells and used western blotting (Figs. 6a and b) and immunofluorescence assay (IFA) (Figs. 6c–f) to detect the expression of DAg proteins. Western blotting detected the expected bands (approximately 22 kDa) only in transfected cells (Figs. 6a and b). Note that a single specific band was detected in lysates prepared from each cell type, suggesting that tgDeV and mmDeV expressed only one DAg isoform. Consistent with the above results, specific signals were observed only in the transfected cells in IFA (Figs. 6c and d, red signals). Together, these data suggest that

the tgDeV and mmDeV initiated replication from the constructed plasmids in the cell culture system.

The DAg proteins predominantly localized to the nucleus 2 days post-transfection (Figs. 6e and f). Interestingly, large viral speckles were observed in the nucleus, similar to those detected in cells infected with HDV (28, 29). These results suggest that tgDeV and mmDeV employ a nuclear replication strategy similar to that used by HDV.

***HBV envelope proteins do not contribute to the production of infectious tgDeV and mmDeV***

As described in the above section "*Candidate helper viruses*", there is no evidence of coinfections of hepadnaviruses with tgDeV or mmDeV. However, this does not necessarily mean hepadnaviruses do not serve as helper viruses for the novel deltaviruses. To determine whether tgDeV or mmDeV utilize the HBV envelope proteins (HBs), we transfected the deltavirus expression plasmids together with an HBs expression vector or the cognate empty vector into Huh7 cells. The culture supernatants were incubated with HepG2-NTCP cells, which are susceptible to HBs-dependent HDV infection (30, 31). HDV served as a control to monitor HBs-dependent virus release and subsequent cell entry. Viral RNA was undetected in supernatants of cells that did not express HBs (Fig. 7a). In contrast, cotransfection of the HBs plasmid released large amounts of HDV RNA into the supernatant, consistent with published data (32), whereas tgDeV or mmDeV RNA was undetectable (Fig. 7a).

We next measured the amounts of viral RNA and detected DAg protein in HepG2-NTCP cells 7 days after incubation with the supernatants. HDV RNA and DAg protein were highly expressed in the infected cells (Fig. 7b and c). HDV infection was

310 inhibited by the preS1 peptide (Myrcludex B), which was shown to inhibit
311 HBs-dependent HDV infection (33). These indicate that HDV infection is indeed
312 mediated by the HBs. On the other hand, tgDeV and mmDeV RNA or DAg protein was
313 undetectable, suggesting that HBs do not contribute to the production of infectious tgDeV
314 or mmDeV.

## Discussion

Important aspects of the evolution of deltaviruses are unknown, such as the origin of HDV and the coevolution of deltaviruses and their helper viruses, mainly because few deltaviruses are known, and they are highly genetically divergent (15-18). Therefore, the resulting large gaps in the deltavirus phylogenetic tree create a formidable obstacle to understanding deltavirus evolution. Here we identified five complete genomes of novel deltaviruses from birds and mammals (Fig. 1 and Supp Fig. 5), which partially fill these phylogenetic gaps (Fig. 5b). Moreover, our present findings reveal that the evolution of deltaviruses is much more complicated than previously thought. For example, one hypothesis states that mammalian deltaviruses codiverged with their mammalian hosts (18). However, our phylogenetic analysis shows that the tree topology of mammalian deltaviruses is incongruent with their hosts'. For example, ovDeV, which we detected in deer, is most closely related to human HDV (Fig. 5b). Further, the distantly related mmDeV and rodent DeV (detected in *Proechimys semispinosus* (18)) were detected in rodent species. These data suggest that deltaviruses were transmitted among mammalian species and did not always codiverge with their hosts. Moreover, we discovered recent interfamily transmission of passerine deltaviruses (tgDeV and its relatives) (Fig. 4e). Therefore, avian and mammalian deltaviruses may have, at least partially, evolved by interspecies transmission.

Our present phylogenetic analysis also gives insights into the origin of human HDVs. As described above, ovDeV and mmDeV are close relatives of human HDVs, suggesting that HDVs arose from other mammalian deltaviruses. Recent studies on the phylogeny of bat deltaviruses support our findings and conclusions (34, 35) (Supp Fig. 7). However, we were unable to exclude the possibility of infection of animal lineages apart from

339 mammals with unknown deltaviruses phylogenetically located between those viral
340 lineages. Further investigations are required for a better understanding of the deltavirus
341 evolution.

342 There is a paucity of knowledge about helper viruses for non-HDV deltaviruses,
343 other than the snake deltavirus (19), although evidence indicates that hepadnaviruses may
344 not serve as helper viruses for the novel deltaviruses discovered here, as suggested for
345 other non-HDV deltaviruses (15-18). Here we only detected bornavirus, circovirus, and
346 polyomavirus, but not hepadnavirus sequences in association with deltavirus-positive
347 SRA data (Table 3). Further, mmDeV was detected in two woodchuck individuals that
348 were demonstrated to be negative for WHV (Table 2 and Supp Table 3). Moreover, we
349 found that HBs did not contribute to the formation of infectious tgDeV and mmDeV in
350 cell culture experiments. These observations suggest that hepadnaviruses do not serve as
351 helper viruses for the novel non-HDV deltaviruses detected here. Further, we were
352 unable to demonstrate that the deltaviruses identified here express proteins similar to the
353 L-HDAg protein (Figs. 6a and b), which is expressed via RNA-editing and is essential for
354 HDV to interact with HBs (2, 37). Although RNA-editing may alter the stop codon of
355 ovDeV DAg, this does not lead to the expression of a large isoform of DAg (L-DAg)
356 protein containing a farnesylation site (Figs. 2b and c). The lack of L-DAg expression
357 was also observed in rodent deltaviruses (18). Therefore, L-DAg expression phenotype
358 may have been acquired after the divergence of ovDeV and the HDV lineage (Fig. 5b).

359 Among the coexisting viruses, only bornavirus produces an envelope glycoprotein
360 (G protein), which might be used by non-HDV deltaviruses to produce infectious virions.
361 Indeed, snake deltavirus utilizes the envelope proteins of reptarenaviruses and
362 hartmaniviruses to produce infectious particles (19). Furthre, HDV forms infectious

363 virions using envelope proteins *in vitro* of RNA viruses such as vesiculovirus and
364 hepacivirus (38). Therefore, the bornavirus G protein might envelop non-HDV
365 deltaviruses.

366 In contrast, the coexisting viruses, circoviruses and polyomaviruses, are
367 nonenveloped. Therefore, it is unlikely that these viruses can serve as helper viruses for
368 the deltaviruses. However, we cannot exclude the possibility that these viral capsid
369 proteins might contribute to the transmissibility of deltaviruses through unknown
370 mechanisms. Additionally, virus-derived sequences in host genomes, called endogenous
371 viral elements (EVEs), might mediate the formation of infectious particles. Here detected
372 the expression of retroviral EVEs in certain deltavirus-positive SRA data (data not
373 shown). Although HDVs do not use retroviral envelope proteins (38), non-HDV
374 deltaviruses might utilize strategies distinct from those employed by HDVs. Alternatively,
375 non-HDV deltaviruses may not require helper viruses and utilize extracellular vesicles for
376 transmission. Further biological experiments, together with molecular surveillance, are
377 therefore required to understand the satellite-helper relationships of deltaviruses.

378 Here we show that the sequences of tgDeV and tgDeV-like viruses, such as pvDeV
379 and lsDeV, are relatively closely related to known vertebrate deltaviruses (Fig. 5b).
380 Although a previous study found a deltavirus from duck, this duck-associated virus was
381 detected in oropharyngeal/cloacal swabs and is distantly related to vertebrate deltaviruses,
382 suggesting the possibility of its dietary origin (15, 18). This may be true for scDeV
383 studied here. scDeV was detected in skin (Table 1). Although scDeV was excluded from
384 our phylogenetic analysis, the amino acid identities between the DAg protein of scDeV
385 with other vertebrate deltaviruses range from 32.7%–39.5% (Fig. 5a). Therefore, scDeV
386 may be derived from contaminants, which should be addressed in the future. In contrast,

tgDeV and tgDeV-like viruses were detected in tissues such as the spleen and muscles (Table 2), suggesting that tgDeV and tgDeV-like viruses are authentic avian deltaviruses.

Here we show that certain novel deltaviruses are transmitted among animal populations (Table 2 and Supp Table 3). Note that few reads were mapped to the virus genomes in some SRA data, which may be attributed to index hopping (39-43) from SRA data containing numerous deltavirus-derived reads. Therefore, these data should be interpreted with caution. Nevertheless, our conclusions are not affected, because they are supported by robust data (Table 2). For data in which index hopping has possibly occurred, further analyses are needed to confirm deltavirus infections.

Our present analysis provides virological insights into important characteristics of deltavirus infections, such as tissue and host tropism. For example, infections with tgDeV (and tgDeV-like viruses), mmDeV, and ovDeV were not limited to the liver and were detected in at least two different tissues (Table 2). These observations are consistent with those of previous studies that non-HDV deltaviruses in multiple organs and blood and that they replicate in numerous cell types (18, 19). Therefore non-HDV deltaviruses may infect diverse tissues and cause systemic infection, viremia, or both. Furthre, tgDeV and mmDeV replicated in human and woodchuck cells (Fig. 6), which is consistent with the ability of the snake deltavirus to replicate in mammalian cells (19). These observations suggest that the host range of deltaviruses is broad and that the helper viruses of non-HDV deltaviruses may be the determinants of host range.

Our analyses further suggest that tgDeV and mmDeV are sensitive to host immune responses. We cross-referenced tgDeV reads and metadata and made an intriguing finding that may contribute to the virus-host interaction. BioProject PRJNA297576 contains 12 RNA-seq data for six zebra finches (44). Interestingly, the tgDeV reads were

411 almost exclusively detected in birds treated with testosterone vs the controls (Supp Fig.
412 8a and Supp Table 5). Therefore, the immunosuppressive effects of testosterone (45) may
413 increase the transcription or replication of tgDeV, or both, to enable detection using
414 RNA-seq. Furthre, when we cross-referenced the mmDeV reads and the metadata, we
415 found that 18 of 20 mmDeV-positive SRA data derived from five individuals were
416 acquired through an experiment lasting 27 weeks (PRJNA291589) (25). Among 18 SRA
417 data, those of one individual (ID 1008) provide insights into mmDeV infection as
418 follows: At first (–3 weeks), mapped reads were not detected, although the proportion of
419 mapped reads were highest at one week and then drastically decreased (Supp Fig. 8b and
420 Supp Table 6). Interestingly, a previous study suggested that the host's immune response
421 can clear rodent deltaviruses (18). Our present observations together with this previous
422 finding, suggest that the host immune response suppresses and then clear deltavirus
423 infections. This may explain the low prevalence of RT-PCR-positive samples of
424 woodchucks and passerine birds as described in the Results (sections "*Circulation of*
425 *tgDeV and tgDeV-like viruses among passerine birds*" and "*Circulation of mmDeV in*
426 *woodchucks*"). Note that latent or low levels of persistent deltavirus infections may occur.
427 Indeed, snake deltavirus establishes a persistent infection in a cell culture system (19).
428 Therefore, deltaviruses might persistently infect host cells with a low level of virus
429 replication, and some stimulations, such as immunosuppression, may trigger robust virus
430 replication. Further studies are therefore required to understand the interactions between
431 deltaviruses and their hosts.

432 Together, our present data contribute to a deeper understanding of the evolution of
433 deltaviruses and suggest the presence of undiscovered deltaviruses that infect diverse
434 animal species. Further investigations will provide further insights into deltavirus

435 evolution.

436

## Materials and methods

*Detection of deltaviruses from publicly available transcriptome data*

Paired-end, RNA-seq data from birds and mammals were downloaded from NCBI SRA (46). The SRA accession numbers used in this study are listed in Supplementary material. The downloaded SRA files were dumped using pfastq-dump (DOI: 10.5281/zenodo.2590842; https://github.com/inutano/pfastq-dump), and then preprocessed using fastp 0.20.0 (47). If genome data of either the same species or the same genus were available, the preprocessed reads were mapped to the corresponding genome sequences (the genome information is available upon request) by HISAT2 2.1.0 (48), and then unmapped paired-end reads were extracted using SAMtools 1.9 (49) and Picard 2.20.4 (http://broadinstitute.github.io/picard/). The extracted unmapped reads were used for *de novo* assembly. If genome data were unavailable, the preprocessed reads were directly used for *de novo* assembly. *De novo* assembly was conducted using SPAdes (50) and/or metaSPAdes (51) 3.13.0 with k-mer of 21, 33, 55, 77, and 99. The resultant contigs were clustered by cd-hit-est 4.8.1 (52, 53) with a threshold of 0.95. Finally, the clustered contigs ≥ 500 nt were extracted by SeqKit 0.9.0 (54), and they were used for the downstream analyses.

Two-step sequence similarity searches were performed to identify RNA virus-like sequences. First, BLASTx searches were performed against a custom database for RNA viruses using the obtained contigs as queries employing BLAST+ 2.9.0 (55) with the following options: -word_size 2, -evalue $1e^{-3}$, max_target_seqs 1. The custom database of RNA viruses consisted of clustered sequences (by cd-hit 4.8.1 with a threshold of 0.98) from viruses of the realm *Riboviria* in the NCBI GenBank (the sequences were downloaded on June 2, 2019) (46). Next, the query sequences with viral hits were

subjected to second BLASTx analyses, which were performed against the NCBI nr database. Finally, the second blast hits with the best hit against deltaviruses were regarded as deltavirus-like agents, and they were used for detailed analyses.

*Confirmation of circularities of deltavirus contigs*

Self dot-plot analyses of linear deltavirus contigs were conducted using the YASS online web server (29). Based on the analysis, the contigs were manually circularized using Geneious 11.1.5 (https://www.geneious.com). Further confirmation of the circularities of deltavirus contigs was obtained by mapping short reads to circular deltavirus contigs using Geneious software as follows. The reads used for the *de novo* assembly were first imported to Geneious, after which they were mapped to the circular contigs using the Geneious mapper. The mapped reads across the circularized boundaries were confirmed manually.

*Detection of possible RNA-editing sites at stop codons of DAg genes*

We used the mapped reads obtained by the analyses described above to detect possible RNA-editing at the stop codons of DAg genes of deltaviruses. We analyzed the nucleotide variations (presence of variations, variant nucleotide(s), and variant frequency) of mapped reads at each of the stop codons of newly identified deltaviruses using the “Find Variation/SNPs” function in Geneious. We used a custom Python script to visualize base quality scores of the NGS reads mapped at the second nucleotide of the stop codon genome. The codes are available at following URL: https://github.com/shohei-kojima/iwamoto_et_al_2020.

*Sequence characterization*

DAg ORFs were detected by the "Find ORFs" function in Geneious with a threshold of 500 nucleotides. Poly(A) signals were manually detected. Putative ribozyme sequences were identified using nucleotide sequence alignment with other deltaviruses. Ribozyme structures were first inferred using the TT2NE webserver (56), and the obtained data were then visualized using the PsudoViewer3 web server (57). We used the visualized data as guides to draw ribozyme structures.

The self-complementarities of deltavirus-like contigs were analyzed using the Mfold web server (58). Coiled-coil domains and nuclear export signals (NLSs) were predicted using DeepCoil (59) and NLS mapper (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi) web servers, respectively.

*Short read mapping for detection of deltavirus infection*

To detect deltavirus-derived reads in publicly available RNA-seq data, short reads were mapped to deltavirus genomes and then the numbers of mapped reads were counted as follows. SRA files were downloaded from NCBI, dumped, and preprocessed following the procedure described above. The preprocessed reads were then mapped to linearized deltavirus contigs by HISAT2 with the default setting. SAM tools were used to extract the mapped BAM files from the resultant BAM files, and the mapped read numbers were counted using BamTools 2.5.1 (60).

*Recovery of a deltavirus genome from RNA-seq data of Pardaliparus venustulus*

Mapped reads obtained from SRR7244693, SRR7244695, SRR7244696, SRR7244697,

and SRR7244698 in the above analysis (section *Short read mapping for detection of deltavirus infection*) were extracted by Geneious. All the extracted reads were co-assembled using Geneious Assembler with the circular contig assembly function. The obtained circular contigs were characterized as described previously.

*Animals and samples*

Zebra finches (n = 30) and Bengalese finches (n = 5) were obtained from breeding colonies at Wada lab, Hokkaido University. The founder birds were originally obtained from local breeders in Japan. Five to twelve birds were kept together in cages in an aviary and were exposed to a 13:11 light-dark cycle. Blood samples were collected from the wing vein using 30 G × 8 mm syringe needles (Becton Dickinson; Franklin Lakes, NJ, USA). Each blood sample was diluted 1.5 times with PBS, frozen immediately on dry ice after collection, and maintained at −80°C until further requirement. These experiments were conducted under the guidelines and with the approval of the Committee on Animal Experiments of Hokkaido University. These guidelines are based on the national regulations for animal welfare in Japan (Law for the Humane Treatment and Management of Animals with partial amendment No.105, 2011).

Woodchucks (*Marmota monax*) were purchased from Northeastern Wildlife (Harrison, ID, USA) and kept at the Laboratory Animal Center, National Taiwan University College of Medicine. At three days of age, the animal supplier infected captive-born woodchucks with WHV from the same infectious pool. Wild-caught woodchucks were infected naturally and live trapped. Serum samples were collected from the woodchucks periodically via the femoral vein by means of venipuncture. Liver tissues of woodchucks were obtained at autopsy, snap-frozen in liquid nitrogen, and

stored at −80°C until RNA extraction. This study used liver tissues from 10 wild-caught and 33 captive-born woodchucks and serum samples from 33 wild-caught and five captive-born woodchucks. In this study, all the experimental procedures involving woodchucks were performed under protocols approved by the Institutional Animal Care and Use Committee of National Taiwan University College of Medicine.

*Real-time and Endpoint RT-PCR detection of deltaviruses from animal specimens*

Total RNAs were isolated from the whole blood samples from zebra finches and serum samples from woodchucks using Quick RNA Viral Kit (Zymo Research; Irvine, CA, USA). The obtained RNA samples were stored at −80°C until further requirement. Total RNAs were also extracted from 50 mg of the woodchuck liver tissues using either Trizol (Thermo Fisher Scientific; Waltham, MA, USA) or ToTALLY RNA kit (Thermo Fisher Scientific; Waltham, MA, USA ) according to the manufacturers' instructions.

The obtained RNA was reverse-transcribed into cDNA using ReverTra Ace qPCR RT Master Mix (TOYOBO; Osaka, Japan), and these were used as templates for real-time PCR analyses. Real-time PCR was performed with KOD SYBR qPCR Mix (TOYOBO) and primers (Supp Table 7) using the CFX Connect Real-Time PCR Detection System (Bio-Rad; Hercules, CA, USA) according to the manufacturer's instructions. The real-time PCR systems for mmDeV and tgDeV were validated using pcDNA3-mmDeV(-) and pcDNA3-tgDeV(-) monomer, respectively, as controls.

End-point RT-PCR was also performed to confirm deltavirus infections. PCR was performed with Phusion Hot Start II DNA Polymerase (Thermo Fisher Scientific) using the above-described cDNAs and primers listed in Supp Table 7. The PCR products were analyzed by agarose gel electrophoresis. The obtained PCR products were purified and

sequenced by Sangar sequencing in FASMAC (Atsugi, Japan).

*Determination of a full genome sequence of deltavirus in passerine birds*

To determine the full genome sequence of detected deltaviruses, the cDNA obtained in the section "*Realtime and Endpoint RT-PCR detection of deltaviruses from animal specimens*" was amplified using illustra GenomiPhi V2 Kit (GE healthcare; Chicago, IL, USA). The amplified DNA was then purified with innuPREP PCRpure Kit (Analytik Jena: Jena, Germany). PCR was performed with Phusion Hot Start II DNA Polymerase using the primers listed in Supp Table 7. The PCR products were analyzed using agarose gel electrophoreses. When single bands were observed, the amplicon was purified with innuPREP PCRpure Kit. When several bands were detected, bands of the expected sizes were extracted and purified using Zymoclean Gel DNA Recovery Kit (Zymo Research). The purified amplicons were sequenced in FASMAC (Atsugi, Japan).

*Phylogenetic analysis*

Deduced amino acid sequences of DAg proteins were used to infer the phylogenetic relationship between deltaviruses. Multiple alignment was performed by MAFFT 7.427 using the E-INS-i algorithms (61), and ambiguously aligned regions were then removed using trimAl 1.2rev59 with the --strict option (62). The phylogenetic relationship was inferred by the maximum likelihood method using RAxML Next Generation v. 0.9.0 (63). The LG+G model, which showed the lowest BIC by prottest3 3.4.2 (64), was used. The reliability of the tree was assessed by 1,000 bootstrap resampling using the transfer bootstrap expectation method (65). The alignment file is available in Supporting materials.

581

582 *Detection of co-infected viruses*

583 To identify co-infected viruses in deltavirus–positive SRAs, a three-step BLASTx search
584 was performed. First, BLASTx searches were performed against a custom database,
585 including RefSeq protein sequences from viruses using the assembled contigs (see the
586 subsection *Detection of deltaviruses from publicly available transcriptome data*) as
587 queries. The custom database was prepared as follows. Virus-derived protein sequences
588 in the RefSeq protein database (46) were downloaded on July 17, 2020, and were
589 clustered by cd-hit 4.8.1 (threshold = 0.9). Then, sequences of more than 100 amino acid
590 residues were extracted using SeqKit 0.10.1 and these were used as a BLAST database.
591 The first BLAST hits were extracted, which were used for the second BLASTx analysis.
592 The second BLASTx analysis was performed against the NCBI RefSeq protein database.
593 The BLAST hits with the best hit to viral sequences were extracted and used for the final
594 BLASTx searches. The final BLASTx searches were performed against the NCBI nr
595 database. The BLAST hits with the best hit to viral sequences were extracted and
596 analyzed manually.

597

598 *Cell culture*

599 HepG2-NTCP cells were cultured with Dulbecco's modified Eagle's medium
600 (DMEM)/F-12 + GlutaMax (Thermo Fisher Scientific) supplemented with 10 mM
601 HEPES (Sigma Aldrich; St. Louis, MO, USA), 100 unit/ml penicillin (Meiji; Tokyo,
602 Japan), 100 mg/ml streptomycin (Meiji), 10% FBS (Sigma Aldrich), 5 μg/ml insulin
603 (Wako; Tokyo, Japan) and 400 g/ml G418 (Nacalai tesque). Huh7 and WCH-17 cells
604 were maintained in DMEM (Wako) containing 10% FBS (Sigma Aldrich), 100 unit/ml

605 penicillin (Meiji), 100 mg/ml streptomycin (Meiji), 100 mM nonessential amino acids
606 (Thermo Fisher Scientific), 1 mM sodium pyruvate (Sigma Aldrich), and 10 mM HEPES
607 (Sigma Aldrich).
608
609 *Antibody production*
610 The peptides corresponding to 65 to 78 aa (DSSSPRKRKRGEGG) of tgDeV DAg and
611 174 to 187 aa (ESPYSRRGEGLDIR) of mmDeV DAg conjugated with cysteine at N
612 terminus were synthesized. Each of the peptides was injected into mice, and antisera were
613 obtained from the mice at 42 days after the peptide injections. Each of the antisera was
614 affinity-purified using the corresponding peptide. The whole procedure was performed in
615 SCRUM (Tokyo, Japan).
616
617 *Rescue of mmDeV and tgDeV*
618 The DNA of negative-strand genomes of mmDeV and tgDeV was synthesized in
619 GenScript Japan (Tokyo, Japan). The synthesized DNAs were inserted into the KpnI
620 -XbaI site of the pcDNA3 vector, designated as pcDNA3-mmDeV(-) monomer and
621 pcDNA-tgDeV (-) monomer. In addition, tandem sequences of mmDeV and tgDeV
622 genome were inserted into the pcDNA3 vector, which were named pcDNA3-mmDeV(-)
623 dimer and pcDNA-tgDeV (-) dimer, respectively. To rescue these viruses,
624 pcDNA3-mmDeV(-) dimer or pcDNA-tgDeV (-) dimer was transfected into Huh7 and
625 WCH-17 cells using Lipofectamine 3000 and Lipofectamine 2000 (Thermo Fisher
626 Scientific), respectively, according to the manufacturer's instructions. The transfected
627 cells were cultured for 48 h and were used for western blotting, IFA to verify DAg protein
628 expression.

*Western blotting*

Cells were lysed with SDS sample buffer [100 mM Tris-HCl (pH 6.8) (Sigma Aldrich), 4% SDS (Nippon gene; Tokyo, Japan), 20% glycerol (Nacalai tesque), 10% 2-mercaptoethanol (Wako)]. The cell lysates were subjected to SDS-PAGE and transferred onto polyvinylidene difluoride membranes (Merck Millipore; Darmstadt, Germany). After blocking the membranes with 5% skim milk (Morinaga; Tokyo, Japan), they were reacted with anti-tgDeV DAg, anti-mmDeV DAg, or anti-actin (Sigma Aldrich) antibodies as primary antibodies, followed by reaction with horseradish peroxidase (HRP)–conjugated secondary antibodies (Cell Signaling Technology; Danvers, MA, USA).

*Indirect immunofluorescence assay (IFA)*

The cells were fixed in 4% paraformaldehyde (Wako) and then permeabilized using 0.3% Triton X-100 (MP Biomedicals; Santa Ana, CA, USA). After blocking the cells by incubation in PBS containing 1% bovine serum albumin (BSA) (KAC; Kyoto, Japan), they were treated with the primary antibodies against HDAg, tgDeV DAg, or mmDeV DAg and then incubated with Alexa555-conjugated secondary antibody (Thermo Fisher Scientific), together with DAPI (Nacalai tesque). To detect deltavirus-positive cells, the fluorescence signal was observed using fluorescence microscopy, BZ-X710 (KEYENCE; Osaka, Japan). High magnification examination of the subcellular localization was performed using confocal microscopy, LSM900 (ZEISS; Oberkochen, Germany).

*Deltavirus preparation and infection assay*

HDV was produced from the culture supernatants of Huh7 cells transfected with HDV (pSVLD3) and HBs (pT7HB2.7) expressing plasmid, as described previously (32, 66). tgDeV and mmDeV were also subjected to the same assay. The supernatants of transfected cells were collected at 6, 9, and 12 days post-transfection, and they were then filtrated and concentrated using 0.45-μm filters and Amicon Ultra (Merck Millipore), according to the manufacturer's instructions. The concentrated supernatants were inoculated into HepG2-NTCP cells with 5% PEG8000 (Sigma Aldrich) for 24 h followed by washing to remove free viruses. The inoculated cells were cultured for 6 days and used for the downstream analyses.

**Acknowledgments**

HDV and HBs expression plasmids were kindly provided by Dr. John Taylor (the Fox Chase Cancer) and Dr. Camille Sureau (Institute National de la Transfusion Sanguine). We thank Dr. Keiko Takemoto (Kyoto University) for her kind help with the setting of computer resources. The super-computing resources were provided by Human Genome Center, the Institute of Medical Science, the University of Tokyo, and the NIG supercomputer at ROIS National Institute of Genetics. All the silhouette images except for woodchuck were downloaded from silhouetteAC (http://www.silhouette-ac.com/).

This study was supported by the Hakubi project at Kyoto University (MH); Grant-in-Aid for Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports, and Technology (MEXT) of Japan, Grant Numbers JP19H04839 (SI), JP16H06429 (KT), JP16K21723 (KT), JP16H0643 (KT), JP17H05821 (MH), and JP19H04833 (MH); the Japan Society for the Promotion of Science KAKENHI, Grant Numbers JP19K16672 (MI), JP20J00868 (MI), and

677 JP20H03499 (KW); AMED Grant Numbers JP20jm0210068j0002 (KW) and

678 JP20fk0310114j0004 (KW).

679

680 **Author contributions**

681 MH conceived the study. MI, KW, and MH designed the study. MI conducted cell culture

682 experiments. MH, JK, SK performed *in silico* analyses. YS, YTL, HLW, KW, and MH

683 prepared and analyzed animal specimens. All the authors analyzed and discussed the data.

684 MI and MH wrote the manuscript.

685

686 **Figure legends**

687 **Figure 1. Genome organization of novel deltaviruses.**

688 Genomes of **(a)** tgDeV, mmDeV, and ovDeV (complete genomes) and **(b)** scDeV and
689 egDeV (partial genomes). Annotations (ORF, poly-A signal, and ribozymes) are shown
690 by colored arrow pentagons. The numbers indicate nucleotide positions. **(c)**
691 Self-complementarities of novel deltaviruses. The predicted RNA structures were
692 visualized using the Mfold web server (58). Red, blue, and green arcs indicate G-C, A-U,
693 and G-U pairs, respectively.

694

695 **Figure 2. Amino acid sequence characterization of putative delta antigens of novel**
696 **deltaviruses.**

697 **(a)** Alignment and functional features of the putative S-HDAg and DAgs of
698 representative HDVs and novel deltaviruses. (Putative) functional domains are shown by
699 colored boxes. Me: arginine methylation site, Ac: lysine acetylation site, P: Serine
700 phosphorylation site. **(b)** ovDeV mRNA (upper panel) and a possible A-to-I RNA-editing
701 site (lower panel). Consensus ovDeV-DAg mRNA sequence and mapped read sequence
702 with potential RNA-edited nucleotides (blue boxes). Pink boxes indicate the ORF of
703 ovDeV DAg. **(c)** Deduced amino acid sequences of ovDeV-DAg proteins translated from
704 the viral mRNA with or without RNA-editing. The blue letter shows the possible
705 RNA-editing site.

706

707 **Figure 3. Mapping coverages of original short reads of each contig.**

708 Mapped read graphs of **(a)** tgDeV, **(b)** mmDeV, and **(c)** ovDeV. Lines, arrow pentagons,
709 and arrowheads indicate viral genomes, ribozymes, and poly(A) signals, respectively.

710 The numbers above the graphs show nucleotide positions. The light pink box indicates a
711 low read depth region in the putative transcript of tgDeV.
712
713 **Figure 4. Interfamily transmission of deltaviruses among passerine birds.**
714 **(a–c)** RT-PCR detection of a deltavirus from *Lonchura striata*. **(a)** Plasmid used for the
715 establishment of real-time PCR detection system for tgDeV and **(b)** the tgDeV circular
716 genome. The blue arrows indicate the primers used for endo-point RT-PCR detection. **(c)**
717 Endo-point RT-PCR for detection of the circular deltavirus genome. M, 100-bp ladder
718 marker. **(d)** Pairwise nucleotide identities between deltaviruses detected in passerine
719 birds. **(e)** Phylogenetic tree of passerine birds positive for deltaviruses. Phylogenetic tree
720 of birds and deltavirus infections are indicated. MYA: million years ago.
721
722 **Figure 5. Phylogenetic analysis of deltaviruses.**
723 **(a)** Heat map of pairwise amino acid sequence identities between deltaviruses. **(b)** The
724 phylogenetic tree was inferred by the maximum likelihood method using an amino acid
725 sequence alignment of representative deltaviruses. Known phenotypes (RNA-editing and
726 expression of the large isoform of DAg protein) and helper virus(es) of each virus are
727 shown on the right. Note that the SDeV phenotypes are shown in gray letters, because
728 there is insufficient information, evidence, or both for the RNA-editing and L-DAg
729 expression. The deltaviruses identified in this study are indicated by the blue circles.
730 Bootstrap values >70 are shown. SDeV: snake deltavirus, RDeV: rodent deltavirus.
731
732 **Figure 6. Detection of tgDeV DAg and mmDeV DAg in cells ectopically expressing**
733 **the tgDeV or mmDeV dimer genome.**

734 **(a** and **b)** Western blotting analysis of Huh7 or WCH-17 cells transfected with a tgDeV or
735 mmDeV dimer-sequence expression plasmid. The numbers on the left side of panels
736 indicate the size marker of protein (kDa). **(c–f)** Indirect immunofluorescence analysis of
737 the expression of tgDeV or mmDeV DAg protein. The cells were observed using
738 fluorescent microscopy **(c** and **d)** or a confocal microscopy **(e** and **f)**. Blue; DAPI, Red;
739 tgDeV or mmDeV DAg. Scale bars = 50 μm (c and d) and 5 μm (e and f).
740
741 **Figure 7. No infectious particle of tgDeV and mmDeV was produced by**
742 **supplementation of HBV envelop proteins.**
743 **(a)** Quantification of deltavirus RNAs in culture supernatants. HDV, tgDeV, or mmDeV
744 expression plasmid was transfected with or without plasmid expressing HBV envelope
745 proteins into Huh7 cells. Viral RNA levels in supernatants were quantified using
746 quantitative RT-PCR (n = 3). **(b** and **c)** HepG2-NTCP cells were incubated with the
747 culture supernatants of the transfectants for 24 h in the presence or absence of 500 nM
748 Myrcludex B (MyrB), an inhibitor of HBV envelope-dependent viral entry. The cells
749 were cultured for an additional 6 days, and viral RNA levels and protein exprssion were
750 analyzed using quantitative RT-PCR (n = 3) **(b)** and IFA **(c)**, respectively. The numbers in
751 (**c**) correspond to those of (**b**). Blue, DAPI; Red, HDV; tgDeV, or mmDeV DAg. Scale
752 bar = 50 μm.
753

## Reference


1. Magnius L, Taylor J, Mason WS, Sureau C, Deny P, Norder H, Ictv Report C. 2018. ICTV Virus Taxonomy Profile: Deltavirus. J Gen Virol 99:1565-1566.
2. Wong SK, Lazinski DW. 2002. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. Proc Natl Acad Sci U S A 99:15118-23.
3. Bergmann KF, Gerin JL. 1986. Antigens of hepatitis delta virus in the liver and serum of humans and animals. J Infect Dis 154:702-6.
4. Bonino F, Heermann KH, Rizzetto M, Gerlich WH. 1986. Hepatitis delta virus: protein composition of delta antigen and its hepatitis B virus-derived envelope. J Virol 58:945-50.
5. Taylor JM. 2020. Infection by Hepatitis Delta Virus. Viruses 12.
6. O'Malley B, Lazinski DW. 2005. Roles of carboxyl-terminal and farnesylated residues in the functions of the large hepatitis delta antigen. J Virol 79:1142-53.
7. Kuo MY, Sharmeen L, Dinter-Gottlieb G, Taylor J. 1988. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. J Virol 62:4439-44.
8. Sharmeen L, Kuo MY, Dinter-Gottlieb G, Taylor J. 1988. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. J Virol 62:2674-9.
9. Chen PJ, Kalpana G, Goldberg J, Mason W, Werner B, Gerin J, Taylor J. 1986. Structure and replication of the genome of the hepatitis delta virus. Proc Natl Acad Sci U S A 83:8774-8.
10. Kos A, Dijkema R, Arnberg AC, van der Meide PH, Schellekens H. 1986. The hepatitis delta (delta) virus possesses a circular RNA. Nature 323:558-60.
11. Wang KS, Choo QL, Weiner AJ, Ou JH, Najarian RC, Thayer RM, Mullenbach GT, Denniston KJ, Gerin JL, Houghton M. 1986. Structure, sequence and expression of the hepatitis delta (delta) viral genome. Nature 323:508-14.
12. Rizzetto M. 2015. Hepatitis D Virus: Introduction and Epidemiology. Cold Spring Harb Perspect Med 5:a021576.
13. Littlejohn M, Locarnini S, Yuen L. 2016. Origins and Evolution of Hepatitis B Virus and Hepatitis D Virus. Cold Spring Harb Perspect Med 6:a021360.
14. Rizzetto M. 2016. The adventure of delta. Liver Int 36 Suppl 1:135-40.
15. Wille M, Netter HJ, Littlejohn M, Yuen L, Shi M, Eden JS, Klaassen M, Holmes

EC, Hurt AC. 2018. A Divergent Hepatitis D-Like Agent in Birds. Viruses 10.

16. Chang WS, Pettersson JH, Le Lay C, Shi M, Lo N, Wille M, Eden JS, Holmes EC. 2019. Novel hepatitis D-like agents in vertebrates and invertebrates. Virus Evol 5:vez021.

17. Hetzel U, Szirovicza L, Smura T, Prahauser B, Vapalahti O, Kipar A, Hepojoki J. 2019. Identification of a Novel Deltavirus in Boa Constrictors. mBio 10.

18. Paraskevopoulou S, Pirzer F, Goldmann N, Schmid J, Corman VM, Gottula LT, Schroeder S, Rasche A, Muth D, Drexler JF, Heni AC, Eibner GJ, Page RA, Jones TC, Muller MA, Sommer S, Glebe D, Drosten C. 2020. Mammalian deltavirus without hepadnavirus coinfection in the neotropical rodent Proechimys semispinosus. Proc Natl Acad Sci U S A doi:10.1073/pnas.2006750117.

19. Szirovicza L, Hetzel U, Kipar A, Martinez-Sobrido L, Vapalahti O, Hepojoki J. 2020. Snake Deltavirus Utilizes Envelope Proteins of Different Viruses To Generate Infectious Particles. mBio 11.

20. Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971-2.

21. Chao YC, Tang HS, Hsu CT. 1994. Evolution rate of hepatitis delta virus RNA isolated in Taiwan. J Med Virol 43:397-403.

22. Krushkal J, Li WH. 1995. Substitution rates in hepatitis delta virus. J Mol Evol 41:721-6.

23. Alvarado-Mora MV, Romano CM, Gomes-Gouvea MS, Gutierrez MF, Carrilho FJ, Pinho JR. 2011. Dynamics of hepatitis D (delta) virus genotype 3 in the Amazon region of South America. Infect Genet Evol 11:1462-8.

24. Fletcher SP, Chin DJ, Ji Y, Iniguez AL, Taillon B, Swinney DC, Ravindran P, Cheng DT, Bitter H, Lopatin U, Ma H, Klumpp K, Menne S. 2012. Transcriptomic analysis of the woodchuck model of chronic hepatitis B. Hepatology 56:820-30.

25. Fletcher SP, Chin DJ, Gruenbaum L, Bitter H, Rasmussen E, Ravindran P, Swinney DC, Birzele F, Schmucki R, Lorenz SH, Kopetzki E, Carter J, Triyatni M, Thampi LM, Yang J, AlDeghaither D, Murreddu MG, Cote P, Menne S. 2015. Intrahepatic Transcriptional Signature Associated with Response to Interferon-alpha Treatment in the Woodchuck Model of Chronic Hepatitis B.

PLoS Pathog 11:e1005103.

26. Ryu WS, Bayer M, Taylor J. 1992. Assembly of hepatitis delta virus particles. J Virol 66:2310-5.

27. Gudima S, He Y, Chai N, Bruss V, Urban S, Mason W, Taylor J. 2008. Primary human hepatocytes are susceptible to infection by hepatitis delta virus assembled with envelope proteins of woodchuck hepatitis virus. J Virol 82:7276-83.

28. Zuccola HJ, Rozzelle JE, Lemon SM, Erickson BW, Hogle JM. 1998. Structural basis of the oligomerization of hepatitis delta antigen. Structure 6:821-830.

29. Noe L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res 33:W540-3.

30. Yan H, Zhong G, Xu G, He W, Jing Z, Gao Z, Huang Y, Qi Y, Peng B, Wang H, Fu L, Song M, Chen P, Gao W, Ren B, Sun Y, Cai T, Feng X, Sui J, Li W. 2012. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis B and D virus. Elife 1:e00049.

31. Iwamoto M, Watashi K, Tsukuda S, Aly HH, Fukasawa M, Fujimoto A, Suzuki R, Aizaki H, Ito T, Koiwai O, Kusuhara H, Wakita T. 2014. Evaluation and identification of hepatitis B virus entry inhibitors using HepG2 cells overexpressing a membrane transporter NTCP. Biochem Biophys Res Commun 443:808-13.

32. Sureau C, Guerra B, Lee H. 1994. The middle hepatitis B virus envelope protein is not necessary for infectivity of hepatitis delta virus. J Virol 68:4063-6.

33. Gripon P, Cannie I, Urban S. 2005. Efficient inhibition of hepatitis B virus infection by acylated peptides derived from the large viral surface protein. J Virol 79:1613-22.

34. Bergner LM, Orton RJ, Broos A, Tello C, Becker DJ, Carrera JE, Patel AH, Biek R, Streicker DG. 2020. Diversification of mammalian deltaviruses by host shifting. bioRxiv doi:10.1101/2020.06.17.156745.

35. Edgar RC, Taylor J, Altman T, Barbera P, Meleshko D, Lin V, Lohr D, Novakovsky G, Al-Shayeb B, Banfield JF, Korobeynikov A, Chikhi R, Babaian A. 2020. Petabase-scale sequence alignment catalyses viral discovery. bioRxiv doi:10.1101/2020.08.07.241729.

36. Anonymous. !!! INVALID CITATION !!! (34, 35).

853 37. Chang FL, Chen PJ, Tu SJ, Wang CJ, Chen DS. 1991. The large form of
854 hepatitis delta antigen is crucial for assembly of hepatitis delta virus. Proc Natl
855 Acad Sci U S A 88:8490-4.
856 38. Perez-Vargas J, Amirache F, Boson B, Mialon C, Freitas N, Sureau C, Fusil F,
857 Cosset FL. 2019. Enveloped viruses distinct from HBV induce dissemination of
858 hepatitis D virus in vivo. Nat Commun 10:2098.
859 39. Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies
860 in multiplex sequencing on the Illumina platform. Nucleic Acids Res 40:e3.
861 40. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. 2014. Analysis,
862 optimization and verification of Illumina-generated 16S rRNA gene amplicon
863 surveys. PLoS One 9:e94249.
864 41. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. 2015. deML: robust
865 demultiplexing of Illumina sequences using a likelihood-based approach.
866 Bioinformatics 31:770-2.
867 42. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M,
868 Podar M, Quince C, Hall N. 2016. A comprehensive benchmarking study of
869 protocols and sequencing platforms for 16S rRNA community profiling. BMC
870 Genomics 17:55.
871 43. Wright ES, Vetsigian KH. 2016. Quality filtering of Illumina index reads
872 mitigates sample cross-talk. BMC Genomics 17:876.
873 44. Fuxjager MJ, Lee JH, Chan TM, Bahn JH, Chew JG, Xiao X, Schlinger BA.
874 2016. Research Resource: Hormones, Genes, and Athleticism: Effect of
875 Androgens on the Avian Muscular Transcriptome. Mol Endocrinol 30:254-71.
876 45. Duffy DL, Bentley GE, Drazen DL, Ball GF. 2000. Effects of testosterone on
877 cell-mediated and humoral immunity in non-breeding adult European starlings.
878 Behavioral Ecology 11:654-662.
879 46. Coordinators NR. 2018. Database resources of the National Center for
880 Biotechnology Information. Nucleic Acids Res 46:D8-D13.
881 47. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ
882 preprocessor. Bioinformatics 34:i884-i890.
883 48. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome
884 alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol
885 37:907-915.

886 49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis
887 G, Durbin R, Genome Project Data Processing S. 2009. The Sequence
888 Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.
889 50. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin
890 VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N,
891 Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly
892 algorithm and its applications to single-cell sequencing. J Comput Biol
893 19:455-77.
894 51. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new
895 versatile metagenomic assembler. Genome Res 27:824-834.
896 52. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large
897 sets of protein or nucleotide sequences. Bioinformatics 22:1658-9.
898 53. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the
899 next-generation sequencing data. Bioinformatics 28:3150-2.
900 54. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit
901 for FASTA/Q File Manipulation. PLoS One 11:e0163962.
902 55. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden
903 TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
904 56. Bon M, Orland H. 2011. TT2NE: a novel algorithm to predict RNA secondary
905 structures with pseudoknots. Nucleic Acids Res 39:e93.
906 57. Byun Y, Han K. 2009. PseudoViewer3: generating planar drawings of
907 large-scale RNA structures with pseudoknots. Bioinformatics 25:1435-7.
908 58. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization
909 prediction. Nucleic Acids Res 31:3406-15.
910 59. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. 2019.
911 DeepCoil-a fast and accurate prediction of coiled-coil domains in protein
912 sequences. Bioinformatics 35:2790-2795.
913 60. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011.
914 BamTools: a C++ API and toolkit for analyzing and managing BAM files.
915 Bioinformatics 27:1691-2.
916 61. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software
917 version 7: improvements in performance and usability. Mol Biol Evol
918 30:772-80.

62. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972-3.

63. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. 2019. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35:4453-4455.

64. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164-5.

65. Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Davila Felipe M, De Oliveira T, Gascuel O. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556:452-456.

66. Kuo MY, Chao M, Taylor J. 1989. Initiation of replication of the human hepatitis delta virus genome from cloned DNA: role of delta antigen. J Virol 63:1945-50.

**Table 1. Summary of novel deltaviruses**

| Virus name | | Host species | Tissue | SRA accession | DDBJ accession | Contig length (nt) | GC content (%) | BLASTx best hit | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | Virus name | Accession | Identity (%) |
| Taeniopygia guttata DeV | tgDeV | *Taeniopygia guttata* | Scapulohumeralis caudalis | SRR2545946 | BR001665 | 1706 | 56.6 | Rodent deltavirus | QJD13558 | 63.3 |
| Marmota monax DeV | mmDeV | *Marmota monax* | Liver | SRR2136906 | BR001661 | 1712 | 53.4 | Hepatitis delta virus | AIR77039 | 60.0 |
| Odocoileus virginianus DeV | ovDeV | *Odocoileus virginianus* | Pedicle | SRR4256033 | BR001662 | 1690 | 56.4 | Hepatitis delta virus | AHB60712 | 66.7 |
| Erythrura gouldiae DeV | egDeV | *Erythrura gouldiae* | Skin | SRR7504989 | BR001660 | 596 | 59.4[a)] | Rodent deltavirus | QJD13555 | 63.5 |
| Serinus canaria-associated DeV | scDeV | *Serinus canaria* | Skin | SRR2915371 | BR001664 | 761 | 54.4[a)] | Hepatitis delta virus | AIR77012 | 36.0 |
| Pardaliparus venustulus DeV | pvDeV | *Pardaliparus venustulus* | Lung, Kidney, Cardiac muscle, Flight muscle, Liver | SRR7244693 SRR7244695 SRR7244696 SRR7244697 SRR7244698 | BR001663 | 1708 | 55.8 | Rodent deltavirus | QJD13562 | 62.4 |
| Lonchura striata DeV | lsDeV | *Lonchura striata* var. *domestica* | Blood | - | LC575944 | 1708 | 56.2 | Rodent deltavirus | QJD13555 | 62.9 |

a) GC content of the partial genome sequences.

**Table 2. Detection of deltavirus-derived reads in RNA-seq data.**

| Virus | BioProject/ BioStudy | SRA | Host | | RPM [a] (read per million) | Tissue |
|---|---|---|---|---|---|---|
| | | | Taxonomy | | | |
| | | | Family | Species | | |
| tgDeV | PRJNA297576 | SRR2545943 | Estrildidae | *Taeniopygia guttata* | 10.28 | Pectoralis |
| | | SRR2545944 | | | 1.02 | Scapulohumeralis caudalis |
| | | SRR2545946 | | | 56.73 | Scapulohumeralis caudalis |
| | PRJNA558524 | SRR9899549 | Emberizidae [b] | *Emberiza melanocephala* | 3.11 | Blood |
| | PRJNA470787 | SRR7244693 | Paridae | *Pardaliparus venustulus* | 10.68 | Lung |
| | | SRR7244695 | | | 2.07 | Kidney |
| | | SRR7244696 | | | 2.65 | Cardiac muscle |
| | | SRR7244697 | | | 7.12 | Flight muscle |
| | | SRR7244698 | | | 1.77 | Liver |
| | PRJNA478907 | SRR7504989 | Estrildidae | *Erythrura gouldiae* | 1.07 | Skin |
| mmDeV | PRJNA291589 | SRR2136906 | | *Marmota monax* | 70.86 | Liver |
| | | SRR2136907 | | | 63.08 | Liver |
| | | SRR2136916 | | | 1.02 | Liver |
| | SRP011132 | SRR437934 | | | 46.34 | PBMC |
| | | SRR437938 | | | 19.83 | PBMC |
| ovDeV | PRJNA317745 | SRR4256033 | | Odocoileus virginianus | 180.73 | Pedicle |
| scDeV | PRJNA300534 | SRR2915371 | | *Serinus canaria* | 9.79 | Skin |

The full version of the table is available as Supplementary Table 3.

a) This table only shows the samples with RPM >1.

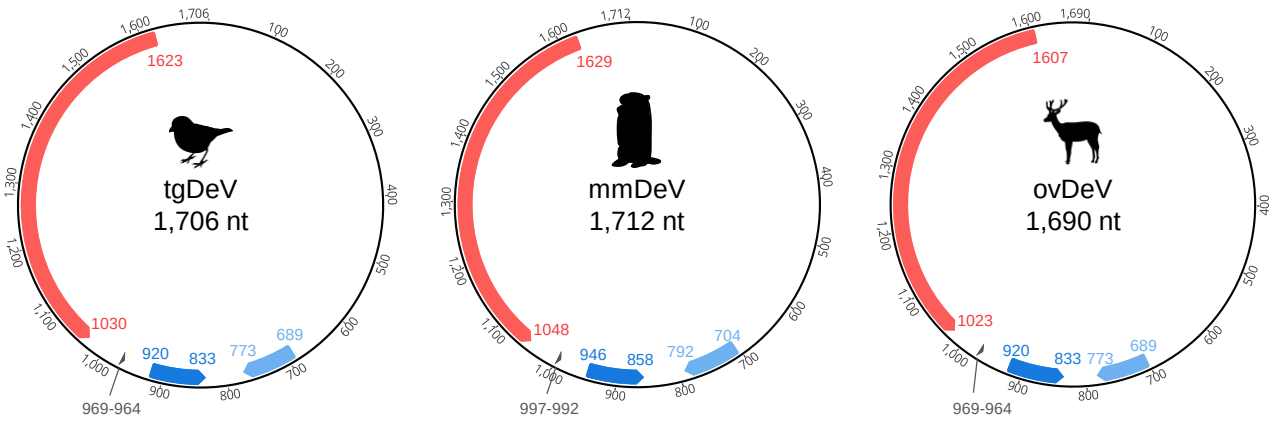
b) Emberizidae is regarded as the subfamily Emberizinae of the family Fringillidae in TimeTree.

**Table 3. Coexisting viruses in deltavirus-positive SRAs**

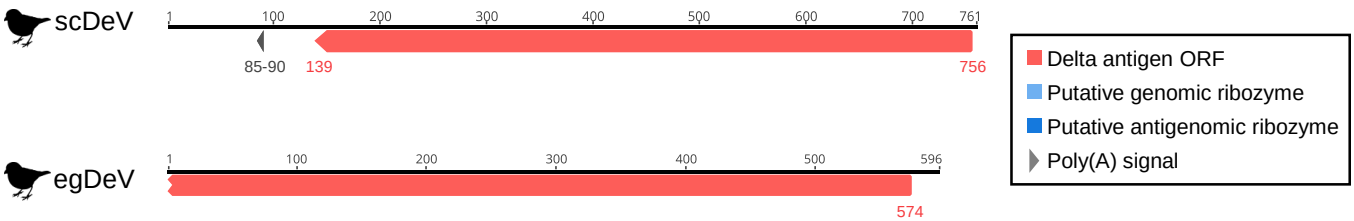
| SRA accession | Host | | Virus name | Envelope | Deltavirus infection |
|---|---|---|---|---|---|
| | Species | Common name | | | |
| SRR2545944 | *Taeniopygia guttata* | Zebra finch | Serinus canaria polyomavirus | - | tgDeV |
| SRR5001849 | *Taeniopygia guttata* | Zebra finch | Serinus canaria polyomavirus | - | tgDeV |
| SRR5001850 | *Taeniopygia guttata* | Zebra finch | Serinus canaria polyomavirus | - | tgDeV |
| SRR5001851 | *Taeniopygia guttata* | Zebra finch | Serinus canaria polyomavirus | - | tgDeV |
| SRR2915371 | *Serinus canaria* | Common canary | Canary bornavirus 3 | + | scDeV |
| | | | Canary circovirus | - | |
| SRR7504989 | *Erythrura gouldiae* | Gouldian finch | Erythrura gouldiae polyomavirus 1 | - | egDeV |

# Fig. 1

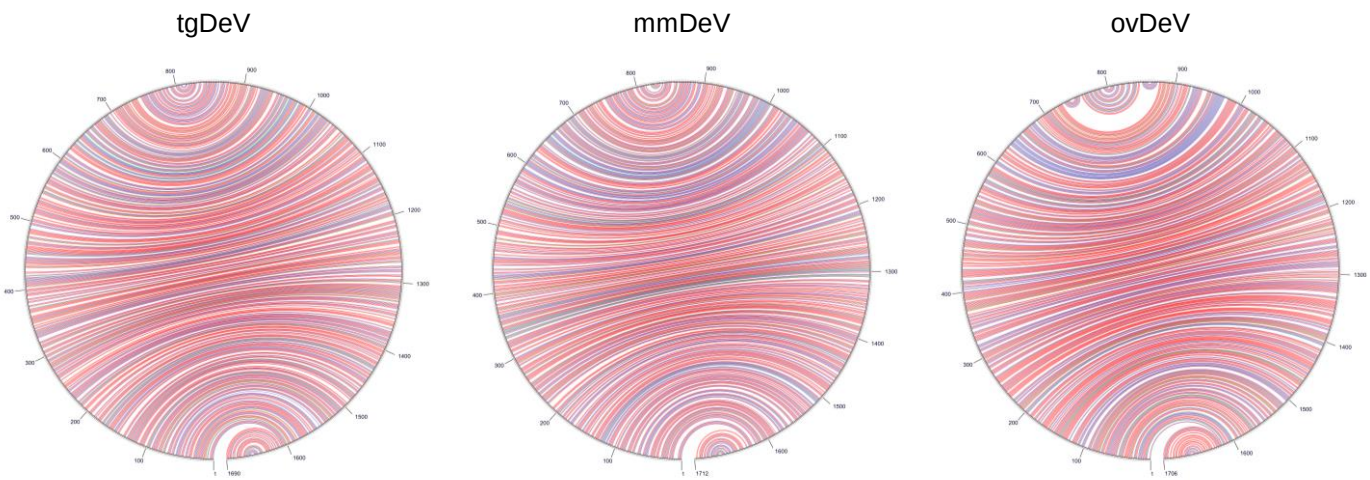
**a**



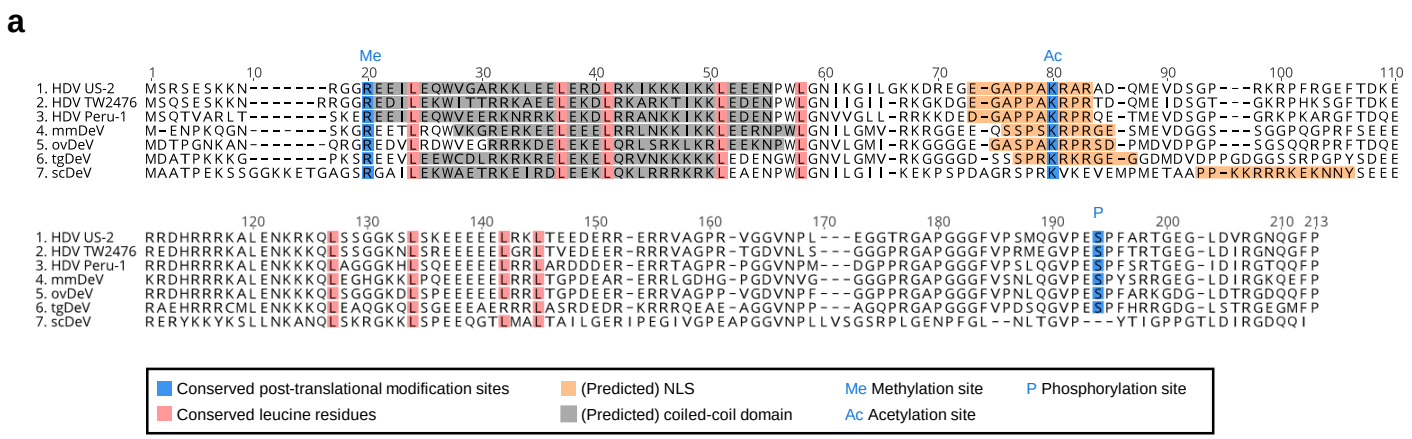

**b**

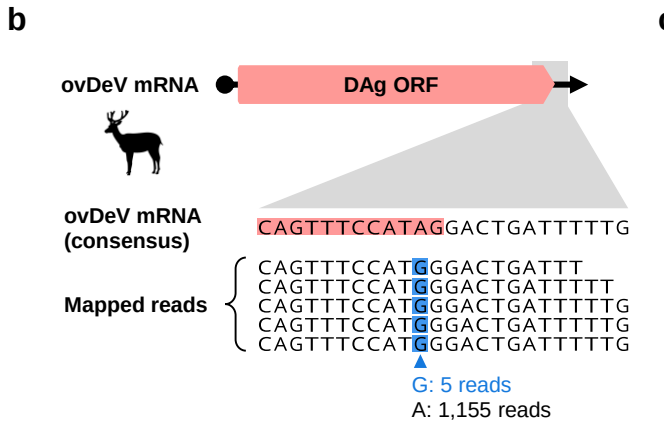


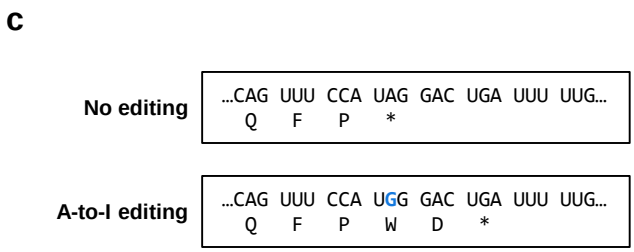

**c**



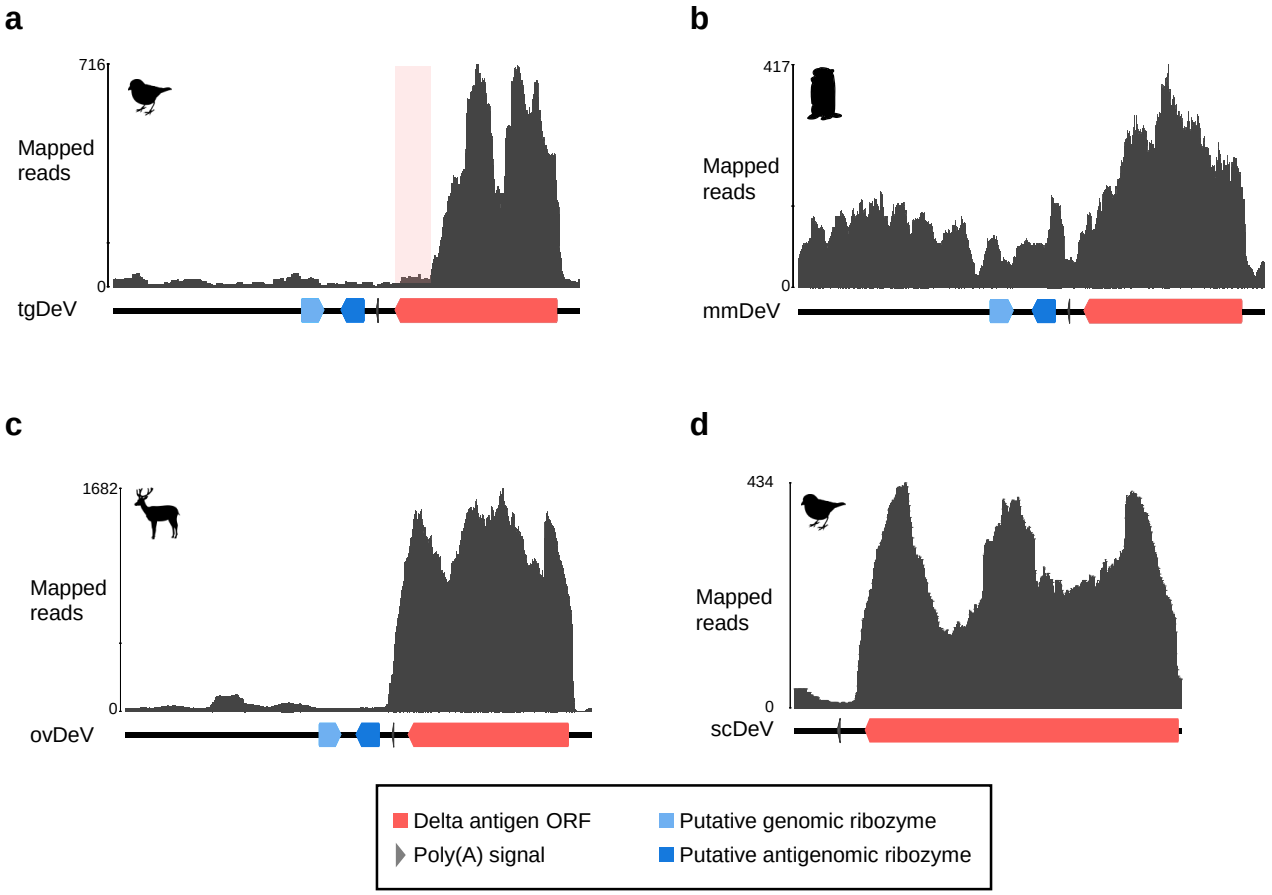
# Fig. 2



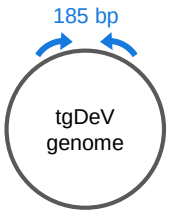



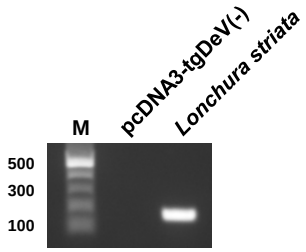



# Fig. 3



# Fig. 4

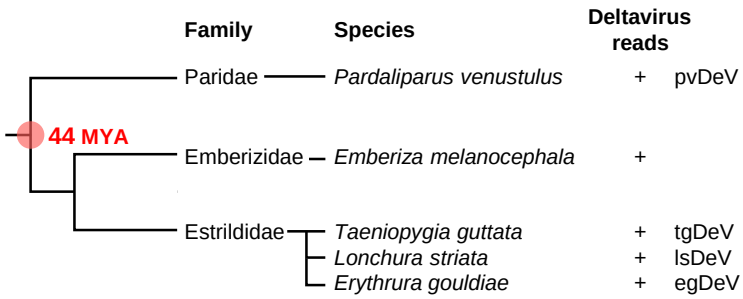




d

**Nucleotide identity (%)**

| | pvDeV | tgDeV | lsDeV |
|---|---|---|---|
| pvDeV | | 98.2 | 98.4 |
| tgDeV | 98.2 | | 98.2 |
| lsDeV | 98.4 | 98.2 | |



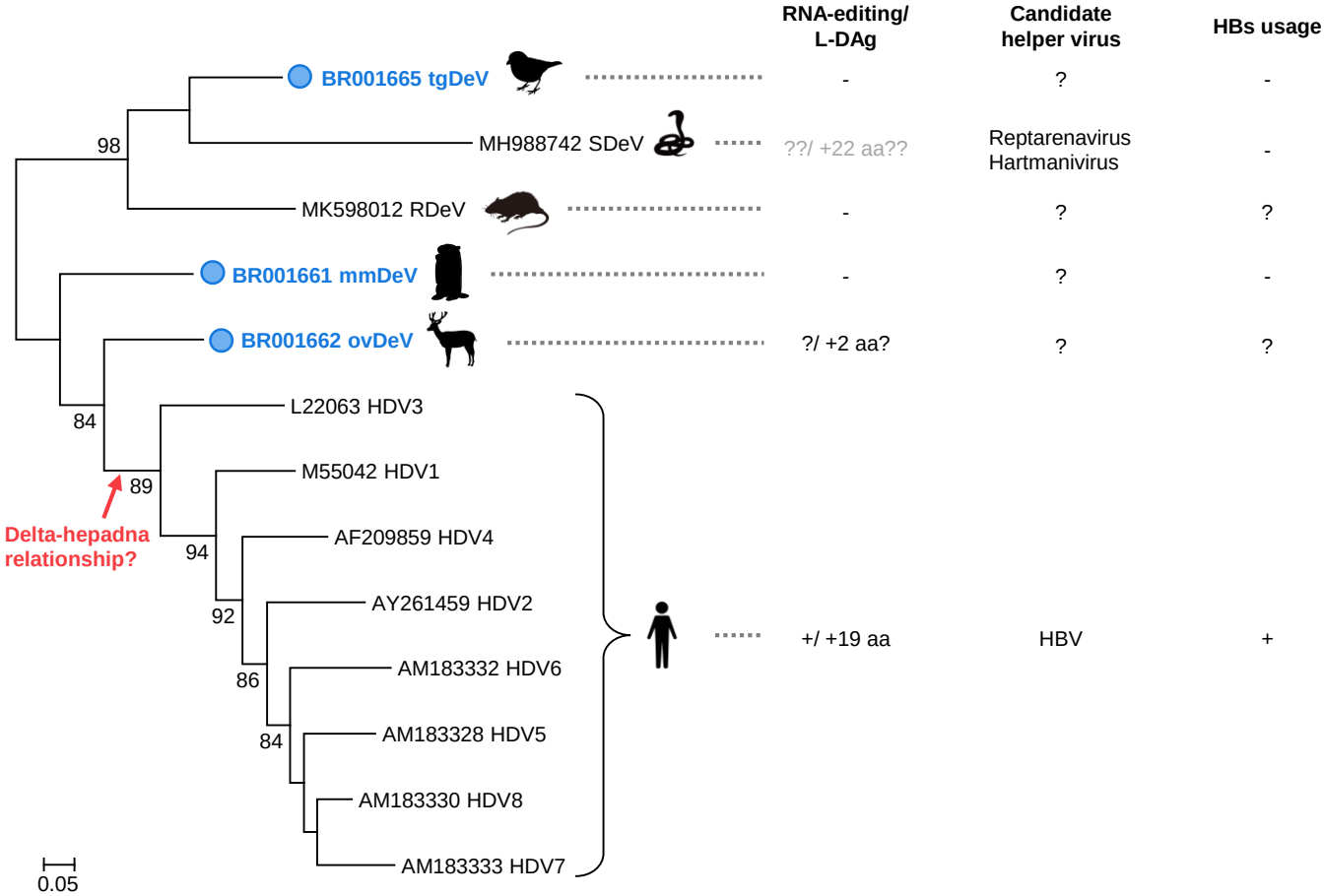
# Fig. 5

**a**

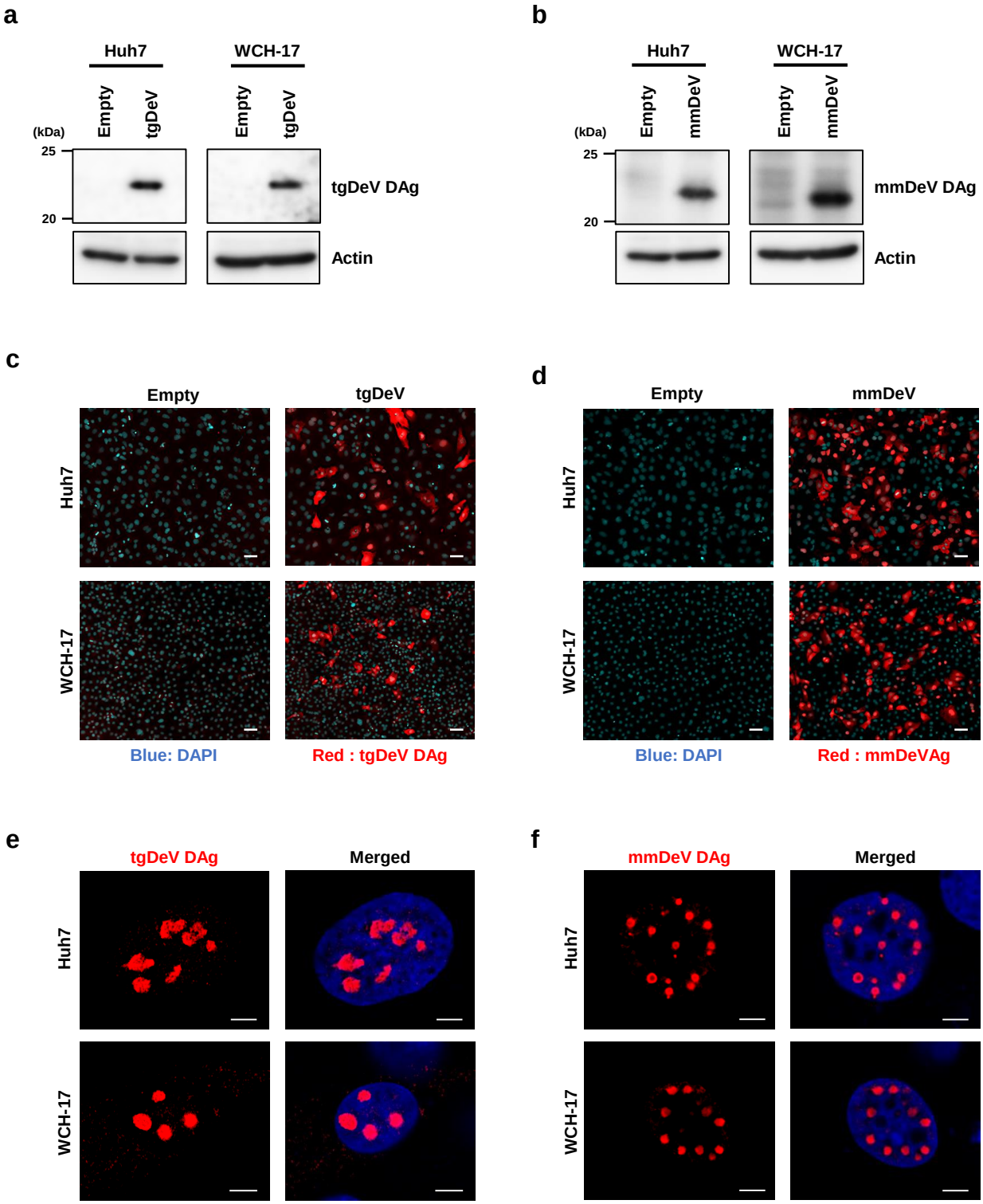
| | M55042 HDV1 | AY261459 HDV2 | L22063 HDV3 | AF209859 HDV4 | AM183328 HDV5 | AM183332 HDV6 | AM183333 HDV7 | AM183330 HDV8 | ovDeV | mmDeV | MK598012 RDeV | tgDeV | MH988742 SDeV | MH824555 duck DeV | scDeV | MN031240 fish DeV | MK962760 toad DeV | MN031239 newt DeV | MK962759 termite DeV |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M55042 HDV1 | | 75.0 | 67.9 | 74.0 | 73.0 | 70.9 | 72.6 | 74.0 | 63.3 | 57.7 | 55.1 | 51.3 | 48.5 | 40.2 | 35.0 | 22.4 | 21.5 | 23.6 | 21.3 |
| AY261459 HDV2 | 75.0 | | 66.2 | 74.9 | 76.5 | 73.3 | 72.4 | 76.9 | 62.6 | 56.4 | 54.8 | 51.5 | 46.2 | 35.6 | 37.1 | 19.0 | 24.6 | 25.6 | 20.4 |
| L22063 HDV3 | 67.9 | 66.2 | | 64.9 | 63.1 | 62.1 | 64.2 | 63.1 | 62.4 | 60.3 | 55.6 | 54.3 | 49.7 | 36.7 | 35.1 | 19.0 | 20.6 | 20.0 | 20.4 |
| AF209859 HDV4 | 74.0 | 74.9 | 64.9 | | 72.4 | 73.3 | 74.0 | 76.9 | 67.0 | 60.8 | 56.1 | 52.8 | 47.2 | 39.4 | 33.2 | 19.0 | 23.6 | 23.3 | 20.4 |
| AM183328 HDV5 | 73.0 | 76.5 | 63.1 | 72.4 | | 74.4 | 74.5 | 79.5 | 62.6 | 53.4 | 52.3 | 50.0 | 45.3 | 36.8 | 38.1 | 20.0 | 23.7 | 24.7 | 22.4 |
| AM183332 HDV6 | 70.9 | 73.3 | 62.1 | 73.3 | 74.4 | | 71.9 | 77.4 | 66.7 | 54.4 | 54.8 | 52.0 | 48.7 | 37.2 | 37.1 | 20.6 | 22.6 | 23.3 | 21.5 |
| AM183333 HDV7 | 72.6 | 72.4 | 64.2 | 74.0 | 74.5 | 71.9 | | 79.1 | 62.7 | 54.5 | 53.4 | 49.6 | 46.4 | 37.9 | 37.7 | 19.1 | 24.2 | 23.4 | 22.0 |
| AM183330 HDV8 | 74.0 | 76.9 | 63.1 | 76.9 | 79.5 | 77.4 | 79.1 | | 65.6 | 57.4 | 54.3 | 52.0 | 46.2 | 37.2 | 39.5 | 19.5 | 22.6 | 27.4 | 20.4 |
| ovDeV | 63.3 | 62.6 | 62.4 | 67.0 | 62.6 | 66.7 | 62.7 | 65.6 | | 67.0 | 57.7 | 58.9 | 51.3 | 37.2 | 36.6 | 24.5 | 23.1 | 22.3 | 21.9 |
| mmDeV | 57.7 | 56.4 | 60.3 | 60.8 | 53.4 | 54.4 | 54.5 | 57.4 | 67.0 | | 54.6 | 55.8 | 49.7 | 39.9 | 32.7 | 22.1 | 23.7 | 22.8 | 21.5 |
| MK598012 RDeV | 55.1 | 54.8 | 55.6 | 56.1 | 52.3 | 54.8 | 53.4 | 54.3 | 57.7 | 54.6 | | 64.1 | 54.7 | 36.0 | 33.0 | 19.9 | 19.4 | 20.0 | 21.8 |
| tgDeV | 51.3 | 51.5 | 54.3 | 52.8 | 50.0 | 52.0 | 49.6 | 52.0 | 58.9 | 55.8 | 64.1 | | 54.2 | 37.2 | 33.8 | 20.2 | 23.3 | 21.8 | 19.9 |
| MH988742 SDeV | 48.5 | 46.2 | 49.7 | 47.2 | 45.3 | 48.7 | 46.4 | 46.2 | 51.3 | 49.7 | 54.7 | 54.2 | | 35.9 | 35.1 | 17.2 | 21.7 | 20.8 | 19.6 |
| MH824555 duck DeV | 40.2 | 35.6 | 36.7 | 39.4 | 36.8 | 37.2 | 37.9 | 37.2 | 37.2 | 39.9 | 36.0 | 37.2 | 35.9 | | 29.5 | 15.3 | 16.8 | 22.3 | 16.2 |
| scDeV | 35.0 | 37.1 | 35.1 | 33.2 | 38.1 | 37.1 | 37.7 | 39.5 | 36.6 | 32.7 | 33.0 | 33.8 | 35.1 | 29.5 | | 17.3 | 21.8 | 22.3 | 17.9 |
| MN031240 fish DeV | 22.4 | 19.0 | 19.0 | 19.0 | 20.0 | 20.6 | 19.1 | 19.5 | 24.5 | 22.1 | 19.9 | 20.2 | 17.2 | 15.3 | 17.3 | | 15.0 | 12.5 | 14.0 |
| MK962760 toad DeV | 21.5 | 24.6 | 20.6 | 23.6 | 23.7 | 22.6 | 24.2 | 22.6 | 23.1 | 23.7 | 19.4 | 23.3 | 21.7 | 16.8 | 21.8 | 15.0 | | 15.4 | 14.0 |
| MN031239 newt DeV | 23.6 | 25.6 | 20.0 | 23.3 | 24.7 | 23.3 | 23.4 | 27.4 | 22.3 | 22.8 | 20.0 | 21.8 | 20.8 | 22.3 | 22.3 | 12.5 | 15.4 | | 15.0 |
| MK962759 termite DeV | 21.3 | 20.4 | 20.4 | 20.4 | 22.4 | 21.5 | 22.0 | 20.4 | 21.9 | 21.5 | 21.8 | 19.9 | 19.6 | 16.2 | 17.9 | 14.0 | 14.0 | 15.0 | |

Identity (%)
High
Low

**b**



# Fig. 6



# Fig. 7