Identifications of novel avian and mammalian deltaviruses filled the large evolutionally gaps and revealed inter-family transmission of deltaviruses

Masashi Iwamoto[1,2], Yukino Shibata[3], Junna Kawasaki[4,5], Shohei Kojima[4,6], Yung-Tsung Li[7], Shingo Iwami[2], 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

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

Abstract

Deltaviruses are unique satellite viruses requiring envelope protein from helper viruses to form infectious particles. Although human hepatitis delta virus (HDV), a satellite virus of a hepadnavirus, has been the only member of the genus *Deltavirus* until 2018, recent studies identified several deltaviruses in vertebrate and invertebrate animals. However, the evolution of deltaviruses, such as the origin of HDV and co-evolution with helper viruses, is still unclear due to the large phylogenetic gaps between the deltaviruses. In this study, by a combination of screening of publicly available RNA-seq datasets and RT-PCR, we identified five full-length, circular genomes of deltaviruses from passerine birds, woodchucks, and white-tailed deer. By mapping pubic RNA-seq data to the newly identified deltaviruses and RT-PCR, we found that the passerine deltaviruses have been circulating among the animal populations, including a bird caught from wild. Interestingly, the passerine deltaviruses show more than 98% nucleotide identities to each other and were detected from birds spanning at least four families in the order Passeriformes, suggesting the recent inter-family transmission events and the potential of broad host ranges of deltaviruses. Phylogenetic analyses showed that the newly identified deltaviruses are relatively closely related to known vertebrate deltaviruses and that newly identified mammalian deltaviruses are the closest relatives of HDV, suggesting the mammalian origin of HDV. Interestingly, there is no evidence that these mammalian deltaviruses express the large isoform of delta antigen that contains a farnesylation site required for HDV to utilize hepadnavirus envelope proteins. Also, there is no evidence of hepadnavirus infections in deltavirus-positive individuals. Therefore, the satellite-helper relationship between HDV and hepadnavirus may have been established relatively recently, after the divergence of the newly

identified non-HDV mammalian deltaviruses and the HDV lineage. Taken together, our findings provide novel insights into the evolution and diversity of deltaviruses and raise the importance of further surveillance for deltaviruses in a wide range of animals.

Introduction

Hepatitis delta virus (HDV), which belongs to the genus *Deltavirus*, has an approximately 1.7-kb circular, negative single-stranded RNA as the virus genome (1, 2). The genome has a single open reading frame (ORF) encoding two viral proteins called small and large hepatitis delta antigens (S- and L-HDAg, approximately 24 and 27 kDa, respectively), which expressed from the same transcription unit via RNA-editing in the stop codon by a host protein, ADAR1 (3-5). The 19-aa extended C-terminal region of L-HDAg protein contains a farnesylation site, which is necessary to interact with envelope proteins from helper viruses (6). The genome structure of HDV is unique in that the genome and antigenome have ribozymes, which are essential for the replication of HDV (7, 8), and show high self-complementarity, resulting in the formation of rod-like shape genome (9-11). Although HDV can alone replicate in the host cells, it needs envelope protein from other "helper" viruses to produce infectious virions. Hepatitis B virus (HBV), which belongs to the family *Hepadnaviridae*, is a well-known helper virus to provide the envelope protein for HDV transmission in human patients (12). Globally, more than 15-20 million people are estimated to be infected by HDV among 350 million people of HBV carriers (14). Coinfection of HDV and HBV accelerate HBV related pathogenesis: severe or fulminant hepatitis and progression to hepatocellular carcinoma than HBV infection alone, although the mechanism of pathogenesis induced by HDV infection has not been fully understood (15).

Since HDV has been the sole member of the genus *Deltaviurs* (2), its evolutionary origin had been enigmatic. However, recent studies discovered several deltaviruses from vertebrates and invertebrate species (16-19), which drastically changed our understanding of the evolution of deltaviruses. These non-HDV deltaviruses are

distantly related to HDV, but apparently share the same origin because they have similar genome structures: their genomes are circular RNA approximately 1.7 kb, encode DAg-like proteins, possess ribozymes sequences, and show high self-complementarities (16-19). These findings would be clues to understanding the origin and evolution of deltaviruses. Indeed, previous studies proposed several interesting hypotheses: for example, a recent study proposed a hypothesis that mammalian deltaviruses have been co-diverged with mammalian species (19). Thus far, however, only a handful of deltaviruses were found, but they are genetically highly diverse (16-19). Therefore, to assess the hypothesis as well as for a deeper understanding of the evolution of deltaviruses, the phylogenetic gaps between the deltaviruses are needed to be filled by exploring novel deltaviruses.

The discoveries of non-HDV deltaviruses have also provided insights into the relationships between deltaviruses and their helper viruses. Recently identified non-HDV deltaviruses apparently did not co-infect with hepadnaviruses, suggesting the presence of other helper viruses (16-19). This is also supported by the lack of large isoform of DAg protein expression, which is mediated by RNA-editing and is necessary for HDV to interact with the hepadnaviral envelope protein as described above, in rodent deltavirus (19). Also, a recent study showed viral envelope proteins from reptarenavirus and hartmanivirus, but not HBV, could render infectivity to snake deltavirus (20). These findings suggest that hepadnaviruses are not helper viruses for non-HDV deltaviruses and that the deltavirus-hepadnavirus relationship has been established only in the HDV lineage. However, the large phylogenetic gap between HDV and a handful of other deltaviruses makes it difficult to assess the hypothesis. Therefore, these observations again raise the importance of exploring deltaviruses and

of experimental investigation for those viruses.

In this study, to understand the evolution of deltaviruses, we explored publicly available transcriptome data from mammals and birds and found novel mammalian and avian deltaviruses. Our phylogenetic analysis suggests that HDVs are apparently originated from non-human mammalian deltaviruses and does not support a co-diversification hypothesis of deltaviruses and mammals. Interestingly, our analyses, together with previous findings, suggest that the satellite-helper relationship between HDV and hepadnavirus may have been established relatively recently, after the divergence of the newly identified non-HDV mammalian deltaviruses and the HDV lineage. Furthermore, we showed evidence of the circulation of the newly identified deltaviruses among animal populations as well as the recent inter-family transmission in birds. Our findings provide novel insights into the evolution as well as virological features of deltaviruses.

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 (21). The SRA accession number 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 by fastp 0.20.0 (22). If genome data of the same species or genus is available, the preprocessed reads were mapped to the corresponding genome sequences by HISAT2 2.1.0 (23), and then unmapped paired-end reads were extracted using SAMtools 1.9 (24) and Picard 2.20.4 (http://broadinstitute.github.io/picard/). The extracted unmapped reads were used for *de novo* assembly. If genome data is unavailable, the preprocessed reads were directly used for *de novo* assembly. *De novo* assembly was performed using SPAdes (25) and/or metaSPAdes (26) 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 (27, 28) with a threshold of 0.95. Finally, the clustered contigs equal to or more than 500 nt were extracted by SeqKit 0.9.0 (29), which 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 using BLAST+ 2.9.0 (30) with the following options: -word_size 2, -evalue 1e-3, max_target_seqs 1. The custom database of RNA viruses was made by clustered sequences (by cd-hit 4.8.1 with a threshold of 0.98) from viruses belonging to the realm *Riboviria* in the NCBI GenBank (the sequences were downloaded on June 2, 2019) (21). The query sequences with viral hits were then subjected to the second BLASTx analyses. The second BLASTx was carried

out against the NCBI nr database. Finally, the second blast hits with the best hit against deltaviruses were regarded as deltavirus-like agents, which were used for detailed analyses.

Confirmation of circularities of deltavirus contigs

Self dot-plot analyses of linear deltavirus contigs were performed in the YASS online web server (31). Based on the analysis, the contigs were manually circularized using Geneious 11.1.5 (https://www.geneious.com). To further confirm circularities of deltavirus contigs, short reads were mapped to circular deltavirus contigs using Geneious software. The reads used for the *de novo* assembly were imported to Geneious, and then mapped to the circular contigs by the Geneious mapper. The mapped reads across the circularized boundaries were manually confirmed.

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

To detect possible RNA-editing at the stop codons of DAg genes of deltaivruses, the mapped reads obtained in the above analyses were used. By the "Find Variation/SNPs" function in Geneious, 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. The quality scores of the variant nucleotides were analyzed manually in Geneious.

Sequence characterization

DAg ORFs was detected by the "Find ORFs" finction in Geneious with a threshold of 500 nucleotides. Poly(A) signals were manually detected. Putative ribozyme sequences

were identified by nucleotide sequence alignment with other deltaviruses. Ribozyme structures were first inferred by TT2NE webserver (32), and then the obtained data were visualized by PsudoViewer3 web server (33). The visualized data were used as guides to draw ribozyme structures.

The self-complementarities of deltavirus-like contigs were analyzed by Mfold web server (34). Coiled-coil domains and nuclear export signals (NLSs) were predicted using DeepCoil (35) and NLSmapper (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 counted the numbers of mapped reads as follows. SRA files were downloaded from NCBI, dumped, and preprocessed as described above. The preprocessed reads were then mapped to linearized deltavirus contigs by HISAT2 with the default setting. The mapped bam files were extracted from the resultant bam files by SAMtools, and the mapped read numbers were counted by BamTools 2.5.1 (36).

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 above.

188

189 *Animals and samples*

190 Zebra finch males (n = 15) and females (n = 15) were obtained from breeding colonies
191 at Wada lab, Hokkaido University. The founder birds were originally obtained from
192 local breeders in Japan. Five to twelve birds were kept together in cages in an aviary on
193 a 13:11 light-dark cycle. Blood samples were collected from the wing vein using 30G ×
194 8 mm syringe needles (Becton Dickinson). Each of the blood samples was diluted 1.5
195 times with PBS, immediately frozen on dry ice after collection, and kept at -80℃ until
196 use. These experiments were conducted under the guidelines and approval of the
197 Committee on Animal Experiments of Hokkaido University. These guidelines are based
198 on the national regulations for animal welfare in Japan (Law for the Humane Treatment
199 and Management of Animals with partial amendment No.105, 2011).

200 Woodchucks (*Marmota monax*) were purchased from Northeastern Wildlife
201 (Harrison, ID, USA), and maintained at the Laboratory Animal Center, National Taiwan
202 University College of Medicine. Captive-born woodchucks were infected with
203 woodchuck hepatitis virus (WHV) at three days of age with the same infectious pool by
204 the animal supplier. Wild-caught woodchucks were infected naturally and live trapped.
205 Serum samples from the woodchucks were collected periodically from the femoral vein
206 using the venipuncture method. Liver tissues of woodchucks were obtained at autopsy,
207 snap-frozen in liquid nitrogen, and stored at −80°C until RNA extraction. Liver tissues
208 from 10 wild-caught and 33 captive-born woodchucks and serum samples from 33 wild-
209 caught and five captive-born woodchucks were used in this study. All experimental
210 procedures involving woodchucks in this study were performed under protocols
211 approved by the Institutional Animal Care and Use Committee of National Taiwan

University College of Medicine.

Realtime 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). The obtained RNA samples were stored at -80°C until use. Total RNAs were also extracted from 50 mg of the woodchuck liver tissues using Trizol (Invitrogen Life Technologies; CA, USA) or ToTALLY RNA kit (Ambion; TX, USA) following the instructions of the manufacturers.

The obtained RNA was reverse-transcribed into cDNA using ReverTra Ace qPCR RT Master Mix (TOYOBO), which 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 CFX Connect Real-Time PCR Detection System (BIO-RAD) according to the manufacturer's instruction. pcDNA3-mmDeV(-) and pcDNA3-rgDeV(-) were used 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. If needed, the obtained PCR products were purified and sequenced by Sangar sequencing in FASMAC (Atsugi, Japan).

Determination of full genome sequence of deltaviruses

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 by illustra GenomiPhi V2 Kit (GE healthcare). The amplified DNA was then purified with innuPREP PCRpure Kit (Analytik Jena). PCR was performed with Phusion Hot Start II DNA Polymerase using the primers listed in Supp Table 7. The PCR products were analyzed by agarose gel electrophoreses. When single bands were observed, the amplicon was purified with innuPREP PCRpure Kit. When there were several bands, 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 delta antigens were used to infer the phylogenetic relationship between deltaviruses. Multiple alignment was performed by MAFFT 7.427 using the E-INS-i algorithms (37), and then ambiguously aligned regions were removed by trimAl 1.2rev59 with the --strict option (38). The phylogenetic relationship was inferred by the maximum likelihood method using RAxML Next Generation v. 0.9.0 (39). The LG+G model, which showed the lowest BIC by prottest3 3.4.2 (40), was used. The reliability of the tree was assessed by 1,000 bootstrap resampling using the transfer bootstrap expectation method (41). The alignment file, as well as the list of accession numbers, are available in Supporting materials.

Detection of co-infected viruses

To identify co-infected viruses in deltavirus-positive SRAs, a three-step BLASTx search was performed. First, BLASTx searches were performed against a custom database, including RefSeq protein sequences from viruses using the assembled contigs (see the

subsection *Detection of deltaviruses from publicly available transcriptome data*) as queries. The custom database was made as follows. All virus-derived protein sequences in the RefSeq protein database (21) were downloaded on July 17, 2020, and were clustered by cd-hit 4.8.1 (threshold = 0.9). Then, sequences of more than 100 amino acid residues were extracted by SeqKit 0.10.1, which were used as a BLAST database. The first BLAST hits were extracted, which were used for the second BLASTx analysis. The second BLASTx analysis was performed against the NCBI RefSeq protein database. The BLAST hits with the best hit to viral sequences were extracted and used for the final BLASTx searches. The final BLASTx searches were performed against the NCBI nr database. The BLAST hits with the best hit to viral sequences were extracted and analyzed manually.

Results

Identification of deltavirus-like sequences in avian and mammalian transcriptome data

To explore deltaviruses, we first assembled 46,359 publicly available RNA-seq data obtained from birds and mammals. Using the resultant contigs as queries, we performed sequence similarity searches, revealing the presences of five deltavirus-like contigs in SRA data from birds and mammals: scapulohumeralis caudalis (muscle) of the zebra finch (*Taeniopygia guttata*), skin of common canary (*Serinus canaria*), skin of Gouldian finch (*Erythrura gouldiae*), liver of Eastern woodchuck (*Marmota monax*), and pedicle of white-tailed deer (*Odocoileus virginianus*) (Table 1), which we herein tentatively named as Taeniopygia guttata deltavirus (tgDeV), Serinus canaria-associated deltavirus (scDeV), Erythrura gouldiae deltavirus (egDeV), Marmota monax deltavirus (mmDeV), and Odocoileus virginianus (ovDeV), respectively. All the contigs showed 36.0-66.7% amino acid sequence similarities to delta antigens (DAg) of known deltaviruses (Table 1, Supp Tables 1 and 2). Among those, tgDeV, mmDeV, and ovDeV contigs are approximately 1,700 nucleotides in length, each of which contains one ORF showing sequence similarity to DAg genes of know deltaviruses (Table 1, Supp Tables 1 and 2). On the other hand, the remaining two contigs, scDeV, and egDeV are 761 and 596 nucleotides in length, respectively (Supp Tables 1 and 2). While the contig obtained from common canary covers the whole ORF region of DAg gene, that of Gouldian finch seems to be partial, which apparently lacks the C-terminal region of DAg (Fig. 1b, Supp Table 2). Note that the nucleotide sequences of tgDeV and egDeV are almost identical (97.7% nucleotide identity), and thus we hereafter mainly analyzed tgDeV instead of both tgDeV and egDeV.

296

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

298 Because the three contigs (tgDeV, mmDeV, and ovDeV) are almost identical in length
299 to full-length genomes of known deltaviruses, we next checked the circularities of the
300 contigs. By self dot-plot analyses, we found each of both ends of these three contigs is
301 identical (Supp Fig. 1), suggesting that the contigs are derived from circular RNA. To
302 further confirm whether these contigs are circular or not, we have mapped the original
303 RNA-seq data to the corresponding circularized contigs using the Geneious mapper. We
304 confirmed that some of the reads properly spanned across the junctions (data not
305 shown), indicating that the contigs are indeed derived from circular RNA. Therefore,
306 the resultant circular contigs of tgDeV, mmDeV, and ovDeV represent the full-length
307 viral genomes, consisting of 1,706, 1,712, and 1,690 nucleotides, respectively (Fig. 1a).
308 Next, we analyzed the nucleotide sequences of the deltaviruses in detail. The GC
309 contents of these deltaviruses are relatively high (53.4-59.4%) (Table 1). We found that
310 poly(A) signal sequences are present downstream of each of the putative ORFs in
311 tgDeV, mmDeV, ovDeV, and scDeV (Figs. 1a and b). We also analyzed the self-
312 complementarity of tgDeV, mmDeV, and ovDeV genomes by mfold web server (34),
313 which indicates that all the genomes show self-complementarities (Fig. 1c).
314 Additionally, we analyzed genomic and antigenomic ribozymes, which are also present
315 or predicted to be present on all the known deltaviruses (16-19). We found putative
316 genomic and antigenomic ribozyme sequences in the three deltaviruses (Figs. 1 and 2,
317 Supp Table 2). All of the genomes have sequences predicted to form typical HDV
318 ribozyme structures (Fig. 2). Overall, these genomic features are identical to those of
319 HDVs and other deltaviruses.

Characterization of DAg proteins encoded by novel deltaviruses

Although recently discovered non-HDV deltaviruses are distantly related to HDVs, DAg proteins of those viruses conserve biochemical features, some biologically relevant amino acid residues, and functional domains, such as post-translational modifications, coiled-coil domains, and nuclear import signals (NLSs) (16-19). Therefore, we also characterized the DAg proteins encoded by newly identified deltaviruses, revealing that many of the features are also conserved among the deltaviruses (Fig. 3a). The isoelectric points of DAg proteins from newly identified deltaviruses are 10.35 to 10.63 (Supp Table 2), which are almost identical to those of known deltaviruses. All the post-translational modification sites in HDVs are also conserved among all the DAg proteins of newly identified deltaviruses except for the serine phosphorylation site on the DAg protein of scDeV. The coiled-coil domain, which is essential for the oligomerization of DAg proteins, is also predicted to be conserved among them. The leucine zipper-like sequences are also present in all the DAg proteins. Furthermore, the NLS is conserved among the DAg proteins except that the predicted NLS of scDeV DAg protein locates at a slightly different position (Fig. 3a).

Human HDV expresses two DAg protein isoforms from one transcription unit by editing the UAG stop codon to UIG (recognized as UGG encoding tryptophan) via the host A-to-I RNA-editing machinery (3). Therefore, we investigated whether the newly identified deltaviruses also use this strategy or not. We mapped short reads from the SRAs, in which the viruses were initially identified, to the corresponding virus genomes, and checked the nucleotide variations at the stop codons. We found that a possible RNA-editing site at the stop codon (UAG) of ovDeV DAg gene. We observed

that there is a 0.4% nucleotide variation (5 of 1160 reads) at the second nucleotide position of the stop codon, all of which are G instead of the consensus nucleotide A (Fig. 3b). The quality scores of the five G variants are 35-41 (Supp Fig. 2), suggesting that this is not caused by sequencing error. Although several possibilities can be considered, such as polymerase errors, this might be produced by A-to-I editing by the ADAR1 as reported for HDV (3). However, this possible RNA editing makes only a 2-aa longer protein due to the presence of a stop codon immediately downstream of the site (Fig. 3c), and the farnesylation motif (CXXQ), which was proposed to be necessary for the interaction with hepadnaviral envelope protein (6), is absent in 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. For tgDeV and mmDeV, we did not find any nucleotide variations of the mapped reads at the stop codons (data not shown).

Deltaviruses were actively infecting in the sampled host animals

To reveal whether the deltaviruses were actively replicating in the sampled animals or not, we evaluated the mapping pattern of viral reads in the above-described mapping analyses. We found that the read depths on the predicted transcribed region (the DAg coding regions to poly-A signals) are much higher than the other parts on the genomes (Fig. 4), indicating that a large part of the viral reads is derived from viral mRNAs. This suggests that the detected deltaviruses were actively transcribing in the sampled host cells. The same trend was also observed in the mapping data to the scDeV genome (Supp Fig. 3).

We noted that the mapping pattern on the tgDeV is slightly different from others.

Although the higher read depth on the DAg ORF region was observed, the read depth covering approximately one-fifth of the ORF is very low (highlighted with a pink box in Fig. 4a). To gain further insights into this observation, we also mapped short reads from another tgDeV positive SRA, which is derived from the same individual as the above analysis but from a different tissue, to the tgDeV genome. We observed the same pattern: the read depth at the downstream region of DAg ORF was lower than that of the upstream region (Supp Fig. 4).

Evidence for circulation and recent inter-family transmission of tgDeV among passerine birds

We next investigated whether the detected deltaviruses have been circulating among animal populations. We first analyzed tgDeV infections in birds. We mapped publicly available RNA-seq data from birds to the tgDeV genome, and then counted the numbers of mapped reads (Table 2 and Supplementary Table 3). In total, we analyzed 6453 SRA data for tgDeV. Among them, surprisingly, tgDeV-derived reads were detected from 34 SRAs from several bird species, such as zebra finch, Gouldian finch, black-headed bunting (*Emberiza melanocephala*), and yellow-bellied tit (*Pardaliparus venustulus*), including the SRAs in which tgDeV and egDeV were initially detected. All of the species belong to the order Passeriformes. Interestingly, tgDeV-positive SRAs were found in nine BioProjects. These nine BioProjects were deposited from independent researchers, and the birds were apparently obtained from different sources. Importantly, the tgDeV-positive sample in SRR9899549 (BioSample accession: SAMN12493457) is derived from a black-headed bunting caught from the wild. These data suggest that tgDeV (or tgDeV-like viruses) has been circulating among a wide range of passerine

392 birds, probably even in wild populations. The tgDeV-positive SRA data were obtained
393 from several tissues, such as blood, kidney, and muscles, suggesting a broad tropism,
394 viremia, and/or systemic infection of tgDeV (see also Discussion).
395 During the above analysis, we found that some of the SRA data contain many
396 mapped reads, which could be enough to reconstruct a full genome of deltavirus.
397 Therefore, we extracted the mapped reads, and then performed *de novo* assembly. We
398 successfully obtained a 1708-nt circular full-genome sequence from yellow-bellied,
399 which was designated as pvDeV. The pvDeV shows 98.2% nucleotide sequence identity
400 to tgDeV, and its DAg protein sequence has the same characteristics as those of tgDeV
401 (Supp Fig. 5).
402 To further investigate deltavirus infections in passerine birds, we tested bird-
403 derived specimens by real-time RT-PCR. We analyzed 30 and 5 whole blood samples
404 from zebra finch and Bengalese finch (*Lonchura striata* var. *domestica*) individuals,
405 respectively. Among them, we found a sample from a Bengalese finch was positive for
406 the real-time RT-PCR analysis. To exclude the possibility of the contamination of a
407 plasmid used for the establishment of real-time PCR system, we also performed RT-
408 PCR using a primer set that can distinguish an amplicon of the viral genome from that
409 of the plasmid (Figs. 5a and b). We obtained a band of the expected size only from the
410 cDNA sample (Fig. 5c), revealing that the bird was indeed positive for tgDeV but not
411 false-positive by contamination. Therefore, we herein named this virus as lsDeV. We
412 further determined the full genome sequence of lsDeV by RT-PCR and sequencing. The
413 full-length genome consists of 1708 nt, showing 98.2% and 98.4% nucleotide identities
414 to tgDeV and pvDeV, respectively. The genome features are almost identical to tgDeV
415 and pvDeV (Supp Fig. 5).

Importantly, the sequence similarities between these passerine deltaviruses are not compatible with the co-divergence hypothesis. According to the TimeTree (42), the divergent time of the passerine deltavirus-positive birds dates back to approximately 44 million years ago (Fig. 6 and Supp Fig. 6). Considering the rapid evolutionary rates of HDVs, which were estimated to be approximately 10^{-3} substitutions per site per year (43-45), these viruses do not seem to have been co-diverged with the hosts, but rather it is most likely that inter-family transmissions have occurred among passerine birds relatively recently.

Evidence for circulation of mmDeV in woodchucks

We next analyzed mmDeV infections using the above mapping strategy using SRA data from animals belonging to the order Rodentia other than mice (*Mus musculus*) and rats (*Rattus norvegicus*). We analyzed 4776 SRA datasets and detected mmDeV reads exclusively from woodchucks, from 20 SRAs derived from 7 individuals (Table 2 and Supplementary Table 3). These mmDeV-positive SRAs were obtained from the same research group, but the animals are apparently obtained in different years (46, 47), suggesting the circulation of mmDeV in the woodchuck populations. The mmDeV-positive 20 SRAs are derived from liver or peripheral blood mononuclear cells (PBMC) samples.

To further investigate mmDeV infections in woodchuck, we tested woodchuck specimens by real-time RT-PCR. We analyzed a total of 81 woodchuck samples (liver, n= 43; serum, n= 38). However, all the samples were negative for mmDeV (data not shown).

No evidence of circulation for other deltaviruses

We also adopted the same strategy to analyze ovDeV and scDeV infections in ruminant animals and passerine birds, respectively (Table 2 and Supp Table 3). We detected ovDeV-derived reads only from five SRAs in which ovDeV was initially identified. The SRA data were obtained from several tissues, such as the brain, muscle, testis, pedicle, and antler, suggesting systemic infection and/or viremia of ovDeV. However, it is not clear whether these samples are derived from multiple individuals or not.

For scDeV, we detected scDeV-derived reads only from the SRA in which the virus was initially detected. Taken together, we could not obtain evidence of the circulation of ovDeV and scDeV.

Phylogenetic analyses on deltaviruses

To infer the evolutionary relationship of deltaviruses, we conducted a phylogenetic analysis using known and the newly identified deltavirus sequences. Here we did not include sequences of recently identified fish, toad, newt, termite, and duck-associated deltaviruses for the phylogenetic analysis because they showed very low amino acid sequence identities to newly identified deltaviruses as well as HDVs (Fig. 7a), which could cause adverse effects on the accuracy of tree (17). We also removed scDeV due to the same reason. Furthermore, we did not use tgDeV-like viruses because the sequence of this virus is almost identical to that of tgDeV. The reconstructed tree showed that the newly identified tgDeV form a strongly supported cluster with snake DeV and rodent DeV, although they are still distantly related to each other (Fig. 7b). mmDeV and ovDeV are more closely related to HDVs than the other deltaviruses.

464 ***Candidate helper viruses for the deltaviruses***

465 To gain insights into helper viruses for the identified deltaviruses, we first analyzed the
466 co-existing viruses in the SRA data in which the deltaviruses were detected. We
467 performed BLASTx analyses using the contigs used to detect deltaviruses. Note that we
468 omitted woodchuck hepatitis virus (WHV) infections in the mmDeV-positive
469 woodchuck-derived SRAs (SRR2136864 to SRR2136999) in this analysis because the
470 woodchucks were experimentally infected with WHV (see below). We also removed
471 viruses infecting non-vertebrate organisms. Our BLASTx analyses revealed that some
472 viruses seemed to have co-existed with the deltaviruses (Table 3 and Supp Table 4). In
473 four of the tgDeV-positive SRAs from zebra finches, we detected Serinus canaria
474 polyomavirus. We also detected canary bornavirus 3 and canary circovirus in the
475 scDeV-positive RNA-seq data. Furthermore, we detected Erythrura gouldiae
476 polyomavirus 1 from the egDeV-positive SRAs. Among these co-existed viruses, only
477 canary bornavirus encode envelope protein. Note that, however, this scDeV and
478 bornavirus-positive SRA is obtained from pooled samples, and thus it is not clear
479 whether scDeV and canary bornavirus 3 infected the same individual or not.

480 We also cross-referenced the presence of mmDeV reads and the metadata, which
481 would also provide insights into the helper virus of mmDeV. Among the 20 mmDeV-
482 positive SRAs, 18 SRAs were obtained from animals experimentally infected with
483 woodchuck hepatitis virus (WHV), which is a hepadnavirus reported to be able to be a
484 helper virus for HDVs (48, 49). However, the other 2 SRAs (SRR437934 and
485 SRR437938) were derived from animals confirmed to be negative for antibodies against
486 WHV as well as WHV DNA (46). This suggests that mmDeV was transmitted to the
487 two animals without WHV, and thus WHV may not have been the helper virus for

488 mmDeV that infected these two individuals.

489

Discussion

The evolution of deltaviruses, such as the origin of HDV and co-evolution of deltaviruses and their helper viruses, has been largely enigmatic. One of the reasons is that only a handful of viruses have been found, but they are genetically highly diverse (16-19). Therefore, large phylogenetic gaps were present between deltaviruses, which made it difficult to understand the evolution of deltaviruses. In this study, we identified five full genomes of novel deltaviruses from birds and mammals (Fig. 1 and Supp Fig. 5), which filled some of the phylogenetic gaps (Fig. 7b). Our analyses revealed that the evolution of deltaviruses is much more complicated than previously thought. A previous study proposed a hypothesis that mammalian deltaviruses might have been co-diverged with their mammalian hosts (19). However, our phylogenetic analysis revealed that the overall tree topology of mammalian deltaviruses is not compatible with that of the host animals: ovDeV, which was detected from deer, is most closely related to human HDV (Fig. 7b). Also, both of mmDeV and rodent DeV (detected from *Proechimys semispinosus* (19)) were detected from rodent species but are distantly related to each other. These data suggest that mammalian deltaviruses have not always been co-diverged with the host animals. Our phylogenetic analysis also gives insights into the origin of human HDVs. As described above, ovDeV and mmDeV are the close relatives of human HDVs, suggesting that HDVs have arisen from mammalian deltaviruses. However, we cannot exclude the possibility that yet-to-be-identified viruses located between these lineages will be found from other lineages of animals.

Helper viruses for non-HDV deltaviruses other than the snake deltavirus (20) are still enigmatic, but hepadnaviruses may not be helper viruses for the newly identified deltaviruses in this study, as suggested for other non-HDV deltaviruses (16-19). Here,

514 we only detected bornavirus, circovirus, and polyomavirus but not hepadnavirus as co-
515 existing viruses in the deltavirus-positive SRAs (Table 3). Additionally, two woodchuck
516 individuals, which were demonstrated to be negative for WHV, were also positive for
517 mmDeV. These observations suggest that hepadnaviruses are not required for the
518 production of infectious virions of non-HDV deltaviruses. Indeed, there is no evidence
519 that the deltaviruses identified in this study express proteins similar to the L-HDAg
520 protein, which is expressed via RNA-editing and is essential for HDV to interact with
521 the HBV envelope protein (3, 50). For tgDeV and mmDeV, there is no evidence of
522 RNA-editing at the stop codons (data not shown). Also, although there might be a
523 possible RNA-editing site at the stop codon of ovDeV (Fig. 3b), this possible RNA-
524 editing does not lead to the expression of a large isoform of DAg (L-DAg) protein
525 containing a farnesylation site due to the presence of stop codon immediately
526 downstream of the editing site (Fig. 3c). Furthermore, the lack of L-DAg was also
527 observed for rodent deltavirus (19). Therefore, the L-HDAg protein expression
528 phenotype may have acquired relatively recently after the divergence of ovDeV and the
529 HDV lineage (Fig. 7b). Considering these observations, the relationship between
530 deltaviruses and hepadnavirus(es) may have been established relatively recently, at least
531 after the divergence of ovDeV and the HDV lineage (Fig. 7b).

532 Among the co-existing viruses, only the bornavirus possesses envelope
533 glycoprotein (G protein), which might be used by deltaviruses for the production of
534 infectious particles of non-HDV deltaviruses. Indeed, snake deltavirus was shown to
535 utilize the envelope proteins of reptarenaviruses and hartmaniviruses to produce
536 infectious particles (20). Also, even HDV was reported to form infectious virions with
537 envelope proteins of several RNA viruses, such as vesiculovirus, flavivirus, and

538 hepacivirus (13). Therefore, the possibility of bornavirus G protein to envelop non-
539 HDV deltaviruses should be addressed in future studies. On the other hand, the other
540 co-existing viruses, a circovirus and polyomaviruses, are non-enveloped, and thus it is
541 unlikely that these viruses can be helper viruses for the deltaviruses. However, we
542 cannot exclude the possibility that these viral capsid proteins might be involved in the
543 transmission of deltaviruses by previously unknown mechanisms, and thus this
544 possibility should also be tested. Another possibility is that virus-derived sequences in
545 the host genomes, called endogenous viral elements (EVEs), might be involved in the
546 formation of infectious particles. We here observed the expression of retroviral EVEs in
547 some deltavirus-positive SRAs (data not shown). Although a previous study revealed
548 that HDVs could not use envelope proteins from retroviruses (13), non-HDV
549 deltaviruses might utilize strategies distinct from HDVs. Alternatively, non-HDV
550 deltaviruses might not need helper viruses for the transmissions. Further biological
551 experiments, together with molecular surveillances, would be necessary to understand
552 the satellite-helper relationships of deltaviruses.

553 In this study, we identified avian deltaviruses relatively closely related to
554 previously identified vertebrate deltaviruses (Fig. 7b). Although a previous study found
555 a deltavirus from duck, this duck-associated virus was detected from
556 oropharyngeal/cloacal swabs and is distantly related to vertebrate deltaviruses,
557 suggesting the possibility of dietary origin (16, 19). This might also be the case for
558 scDeV found in this study. scDeV was detected from the skin (Table 1), which can be
559 easily contaminated by the outside environment. Additionally, although scDeV was not
560 included in the phylogenetic analysis, the amino acid identities of DAg protein of
561 scDeV to the other vertebrate deltaviruses are very low (32.7-39.5% amino acid

562 identities to the DAgs of other vertebrate deltaviruses) (Fig. 7a). Therefore, scDeV
563 might be derived from contaminants, which should be addressed by further studies. On
564 the other hand, tgDeV and tgDeV-like viruses were detected from several tissues, such
565 as blood, kidney, and muscles (Table 2), and are clustered with snake and rodent
566 deltaviruses in the phylogenetic analysis (Fig. 7b). These observations suggest that
567 tgDeV and tgDeV-like viruses are authentic avian deltaviruses.

568 Here we showed the evidence that some of the newly identified deltaviruses have
569 been circulating among animal populations. Interestingly, tgDeV and tgDeV-like viruses
570 were detected from a wide range of passerine birds, suggesting recent inter-family
571 transmission events of these viruses (Fig. 6). Note that in some SRAs, only a few reads
572 could be mapped to the virus sequences, which might be attributed to index hopping
573 (51-55) from SRAs containing many deltavirus-derived reads, and thus numbers of
574 infected individuals and infected species might be overestimated (Table 2 and Supp
575 Table 3). Nonetheless, this does not affect the conclusion because at least some of the
576 SRA data from four species of birds (zebra finch, Gouldian finch, yellow-bellied tit, and
577 black-headed bunting) were independently obtained and contain many viral reads.
578 Interestingly, one of the deltavirus-positive SRAs (SRR9899549) was obtained from a
579 black-headed bunting caught from wild, suggesting the inter-family transmission may
580 have also occurred in wild populations. Taken together, this study provides the first
581 evidence of inter-family transmission of deltavirus, which contributes to a deeper
582 understanding of not only the evolution but also the transmission of deltaviruses.

583 Our analysis would also provide virological insights into the deltaviruses in
584 infected individuals, such as tissue and host tropisms. The infections of tgDeV (and
585 tgDeV-like viruses), mmDeV, and ovDeV were not limited to the liver and detected

586 from more than two different tissues (Table 2). These observations are identical to the
587 previous findings that revealed the presence of deltaviruses in multiple organs and
588 blood, and non-HDV deltaviruses could replicate in many cell types (19, 20). Therefore,
589 non-HDV deltaviruses may be able to infect a wide range of tissues, cause systemic
590 infection, and/or viremia. Additionally, rodent deltavirus and even snake deltavirus were
591 shown to replicate also in human hepatocyte cells (20). These observations suggest that
592 the host tropisms are broad in terms of the host species. Therefore, non-HDV
593 deltaviruses have the potential to replicate in a wide range of host cells and species, and
594 their helper viruses may be the determinant of cell and host tropisms.

595 Our analyses also suggest that tgDeV and mmDeV are sensitive to the host
596 immune responses. We cross-referenced the presence of tgDeV reads and the metadata
597 and found an interesting observation that may be involved in the virus-host interaction.
598 BioProject PRJNA297576 contains twelve RNA-seq data from six zebra finch
599 individuals, among which three individuals were treated with testosterone, and the other
600 three were non-treated control (56). Interestingly, tgDeV reads were almost exclusively
601 detected from birds treated with testosterone (Supp Fig. 7a and Supp Table 5).
602 Testosterone is a hormone known to suppress the immune systems (57). Therefore, the
603 testosterone treatment might have enhanced the transcription and/or replication of
604 tgDeV, and thus the tgDeV reads were prone to be detectable by RNA-seq. We also
605 cross-referenced the presence of mmDeV reads and the metadata. Interestingly, 18 of
606 the 20 mmDeV-positive SRAs, which were derived from 5 individuals, were obtained
607 from a long-term time-course experiment spanning 27 weeks (PRJNA291589) (47).
608 Among the 18 SRAs, data from one individual (ID 1008) would provide insights into
609 the mmDeV infection: at first (-3 week) no mapped read was detected from the

individual, but the proportion of mapped reads in the ID 1008 are highest at one week, and then drastically decreased with time (Supp Fig. 7b and Supp Table 6). Interestingly, a previous study also suggested that the host immune responses can clear rodent deltaviruses (19). Our observations, together with the previous finding, suggest that the host immune responses can suppress deltavirus infections and might be able to clear deltavirus infections. However, we cannot exclude the possibility of latent or low levels of persistent infections. Indeed, snake deltavirus was shown to establish a persistent infection, at least in cell culture (20). Therefore, deltaviruses might persistently infect the host cells with a low level of replications, and some stimulations, such as immunosuppression, might trigger the replications. Further studies are needed to understand the interactions between deltaviruses and their hosts.

The pathogenicities of non-HDV deltaviruses are still unknown. The newly identified deltaviruses in this study were detected from apparently healthy individuals except for woodchucks experimentally infected with WHV. As described above, HDVs are known to make clinical outcomes of HBV infections worse (15). Therefore, non-HDV deltaviruses might also affect the outcomes of co-infected viruses. In this study, we showed that bornavirus, circovirus, and polyomavirus co-existed with the deltaviruses (Table 3). These co-existing viruses are reported to be associated with diseases in avian species: for example, bornaviruses are the etiological agents of proventricular dilatation disease (or -like diseases) in birds (58, 59). Therefore, epidemiological studies of deltaviruses, together with co-existing viruses, would be useful to understand the pathogenicity of deltaviruses.

We observed that the patterns of mapped read depths in tgDeV were different from those in other deltaviruses identified in this study (Fig. 4a and Supp Fig. 4). This might

be because the transcription mechanism of tgDeV is distinct from those of the other deltaviruses. All deltaviruses possess the strictly conserved poly(A) signal sequence (5'-AAUAAA-3') downstream of the DAg ORF. Our analyses clearly showed that the poly(A) signals worked in mmDeV and ovDeV, since the read depth on the whole regions from ORF to poly(A) signal are much higher than the other parts (Fig. 4). On the other hand, the read depth of tgDeV on the downstream region of DAg ORF to poly(A) signal was much lower than the remaining part of ORF, suggesting the possibility that tgDeV has different transcription mechanism. However, there is no alternative poly(A) signal-like sequence near the breakpoint. Therefore, the strange mapping pattern may be derived from artifacts during the library preparation or sequencing. Further molecular biological studies are needed to understand the replication mechanisms of deltaviruses.

Taken together, our findings contribute to a deeper understanding of the evolution as well as the virological features of deltaiviruses. Additionally, our and previous findings of novel deltaviruses also suggest the presence of yet-to-be-identified deltaviruses in diverse eukaryotes. Further investigations on animals would give further insights into the evolution of deltaviruses.

Acknowledgments

We thank Dr. Keiko Takemoto (Kyoto University) for her kind help with the bioinformatic analyses. 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 JP17H05821 (MH), JP19H04833 (MH) ; the Japan Society for the Promotion of Science KAKENHI, Grant Numbers JP19K16672 (MI), JP20J00868 (MI), and JP20H03499 (KW); AMED Grant Numbers JP20jm0210068j0002 (KW) and JP20fk0310114j0004 (KW), and JP20wm0325007h0001 (SI); JST MIRAI (SI).

Figure legends

Figure 1. Genome organizations of newly identified deltaviruses.

Schematic figures for the genome organizations of **(a)** tgDeV, mmDeV, and ovDeV (whole genomes) and **(b)** scDeV and egDeV (partial genomes). Annotations (ORF, poly-A signal, and ribozymes) are shown by colored arrow pentagons. The numbers indicate nucleotide positions. **(c)** Self-complementarities of newly identified deltaviruses. Circular structure plots of tgDeV, mmDeV, and ovDeV are shown. The RNA structures were predicted, and then visualized on the mfold web server (34). Red, blue, and green arcs indicate G-C, A-U, and G-U pairs, respectively.

Figure 2. Putative structures of ribozymes on deltavirus genomes and antigenomes.

Putative genomic and antigenomic ribozyme structures of tgDeV, mmDeV, and ovDeV. Gray numbers show nucleotide positions on each putative ribozyme. The blue letters show the names of secondary structural elements of ribozymes. Catalytic active sites are shown by pink triangles.

Figure 3. Amino acid sequence characterization of putative delta antigens of newly identified deltaviruses.

(a) Amino acid sequence alignment of putative S-HDAg and DAgs of representative HDVs and newly identified deltaviruses, respectively, are shown. (Putative) functional domains are shown by colored boxes. Me: arginine methylation site, Ac: lysine acetylation site, P: Serine phosphorylation site. **(b)** Schematic figure of ovDeV mRNA (upper panel) and a possible A-to-I RNA-editing site on it (lower panel). Consensus ovDeV DAg mRNA sequence and mapped read sequence with possible RNA-edited

nucleotide (blue boxes) are shown. Pink boxes indicate the ORF region of ovDeV DAg. **(c)** Deduced amino acid sequences of ovDeV DAg proteins translated from mRNA with or without RNA-editing. Blue letter shows the possible RNA-editing.

Figure 4. Mapping coverages of original short reads of each of the contigs.
Mapped read graphs of **(a)** tgDeV, **(b)** mmDeV, and **(c)** ovDeV are shown. Lines, arrow pentagons, and arrowheads indicate viral genomes, ribozymes, and poly(A) signals, respectively. Numbers above the graphs show nucleotide positions. Light pink box indicates low read depth region in the putative transcript of tgDeV.

Figure 5. Detection of a delta virus from *Lonchura striata* by RT-PCR
Shematic figure of **(a)** the plasmid used for the establishment of real-time PCR detection system for tgDeV and **(b)** tgDeV circular geonome. The blue arrows indicate the primers used for endo-point RT-PCR detection. **(c)** Endo-point RT-PCR for detection of circular deltavirus genome. M: 100 bp ladder marker.

Figure 6. Interfamily transmission of deltaviruses among passerine birds.
(a) Schematic tree of passerine birds positive for deltaviruses. The tree of birds, presence of deltaviruses and thier names are indicated. MYA: million years ago. **(b)** Pairwise nuclotide identities between deltaviruses detected from passerine birds.

Figure 7. Phylogenetic relationship of deltaviruses.
(a) Heatmap of pairwise amino acid sequence identities between deltaviruses. **(b)** The phylogenetic tree was inferred by the maximum likelihood method using an amino acid

714 sequence alignment of representative deltaviruses. Known phenotypes, RNA-editing
715 and large isoform of DAg protein expression, and helper virus(es) of each virus are
716 shown at the right side of tree. Note that the SDeV phenotypes are shown in gray letters
717 because there is insufficient information and/or evidence for the RNA-editing and L-
718 DAg expression. The deltaviruses identified in this study were shown with blue circles.
719 Bootstrap values more than 70 are shown. SDeV: snake deltavirus, RDeV: rodent
720 deltavirus.

Table 1. Summary of newly identified 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 contents were calculated from partial genome sequence.

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

Full version of the table is avalable as Supplementary Table 3.

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

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

Table 3. Co-exisiting 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

Reference

1. Taylor JM. 2020. Infection by Hepatitis Delta Virus. Viruses 12.
2. 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.
3. 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.
4. 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.
5. 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.
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. Perez-Vargas J, Amirache F, Boson B, Mialon C, Freitas N, Sureau C, Fusil F, Cosset FL. 2019. Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. Nat Commun 10:2098.
14. 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.

15. Rizzetto M. 2016. The adventure of delta. Liver Int 36 Suppl 1:135-40.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. Coordinators NR. 2018. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 46:D8-D13.
22. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884-i890.
23. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907-915.
24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078-9.
25. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455-77.

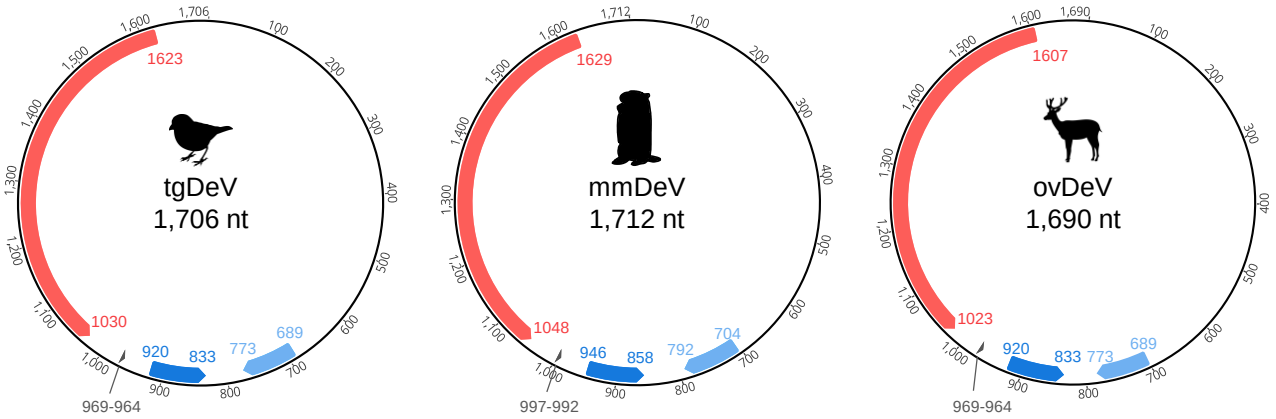
26. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res 27:824-834.
27. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658-9.
28. Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150-2.
29. Shen W, Le S, Li Y, Hu F. 2016. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One 11:e0163962.
30. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.
31. Noe L, Kucherov G. 2005. YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Res 33:W540-3.
32. Bon M, Orland H. 2011. TT2NE: a novel algorithm to predict RNA secondary structures with pseudoknots. Nucleic Acids Res 39:e93.
33. Byun Y, Han K. 2009. PseudoViewer3: generating planar drawings of large-scale RNA structures with pseudoknots. Bioinformatics 25:1435-7.
34. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406-15.
35. Ludwiczak J, Winski A, Szczepaniak K, Alva V, Dunin-Horkawicz S. 2019. DeepCoil-a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics 35:2790-2795.
36. Barnett DW, Garrison EK, Quinlan AR, Stromberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 27:1691-2.
37. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772-80.
38. 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.
39. 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.

40. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164-5.
41. 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.
42. Hedges SB, Dudley J, Kumar S. 2006. TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971-2.
43. Chao YC, Tang HS, Hsu CT. 1994. Evolution rate of hepatitis delta virus RNA isolated in Taiwan. J Med Virol 43:397-403.
44. Krushkal J, Li WH. 1995. Substitution rates in hepatitis delta virus. J Mol Evol 41:721-6.
45. 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.
46. 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.
47. 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.
48. Ryu WS, Bayer M, Taylor J. 1992. Assembly of hepatitis delta virus particles. J Virol 66:2310-5.
49. 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.
50. Chang FL, Chen PJ, Tu SJ, Wang CJ, Chen DS. 1991. The large form of hepatitis delta antigen is crucial for assembly of hepatitis delta virus. Proc Natl Acad Sci U S A 88:8490-4.

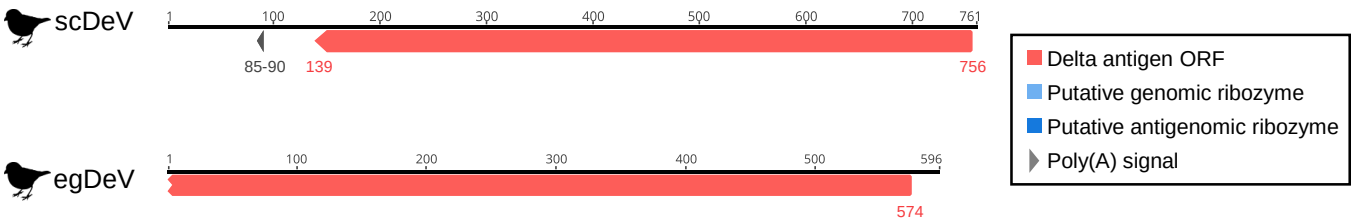
51. Kircher M, Sawyer S, Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Res 40:e3.
52. Nelson MC, Morrison HG, Benjamino J, Grim SL, Graf J. 2014. Analysis, optimization and verification of Illumina-generated 16S rRNA gene amplicon surveys. PLoS One 9:e94249.
53. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J. 2015. deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. Bioinformatics 31:770-2.
54. D'Amore R, Ijaz UZ, Schirmer M, Kenny JG, Gregory R, Darby AC, Shakya M, Podar M, Quince C, Hall N. 2016. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17:55.
55. Wright ES, Vetsigian KH. 2016. Quality filtering of Illumina index reads mitigates sample cross-talk. BMC Genomics 17:876.
56. Fuxjager MJ, Lee JH, Chan TM, Bahn JH, Chew JG, Xiao X, Schlinger BA. 2016. Research Resource: Hormones, Genes, and Athleticism: Effect of Androgens on the Avian Muscular Transcriptome. Mol Endocrinol 30:254-71.
57. Duffy DL, Bentley GE, Drazen DL, Ball GF. 2000. Effects of testosterone on cell-mediated and humoral immunity in non-breeding adult European starlings. Behavioral Ecology 11:654-662.
58. Weissenbock H, Sekulin K, Bakonyi T, Hogler S, Nowotny N. 2009. Novel avian bornavirus in a nonpsittacine species (Canary; Serinus canaria) with enteric ganglioneuritis and encephalitis. J Virol 83:11367-71.
59. Rubbenstroth D, Rinder M, Stein M, Hoper D, Kaspers B, Brosinski K, Horie M, Schmidt V, Legler M, Korbel R, Staeheli P. 2013. Avian bornaviruses are widely distributed in canary birds (Serinus canaria f. domestica). Vet Microbiol 165:287-95.

Fig. 1

a


tgDeV
1,706 nt
1623
1030
920
833
773
689
969-964
mmDeV
1,712 nt
1629
1048
946
858
792
704
997-992
ovDeV
1,690 nt
1607
1023
920
833
773
689
969-964

b


scDeV
85-90
139
756
egDeV
574
Delta antigen ORF
Putative genomic ribozyme
Putative antigenomic ribozyme
Poly(A) signal

c

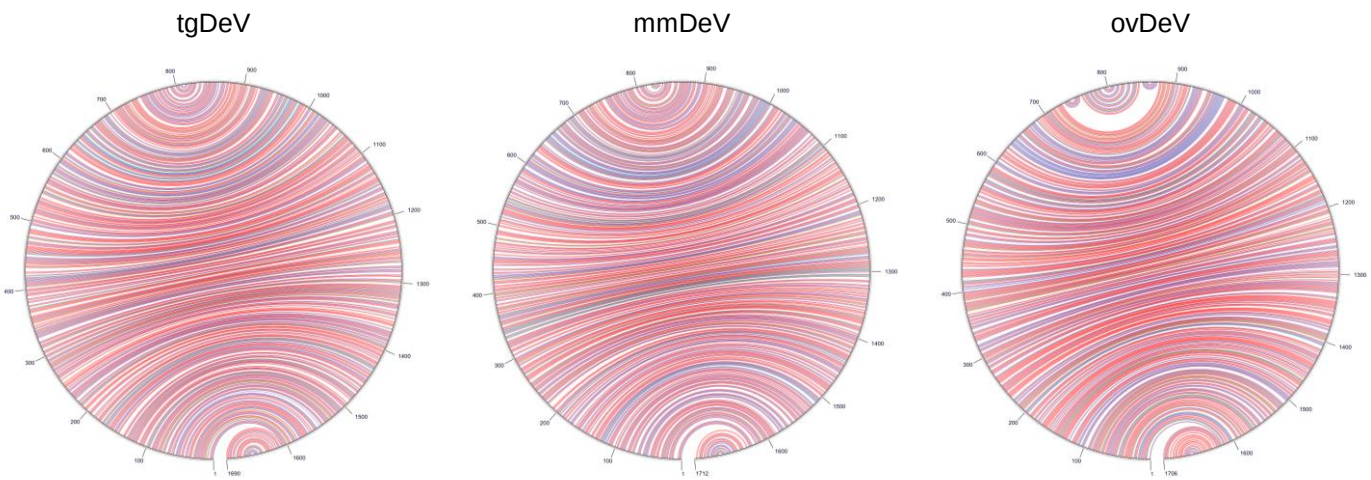

tgDeV
mmDeV
ovDeV

Fig. 2

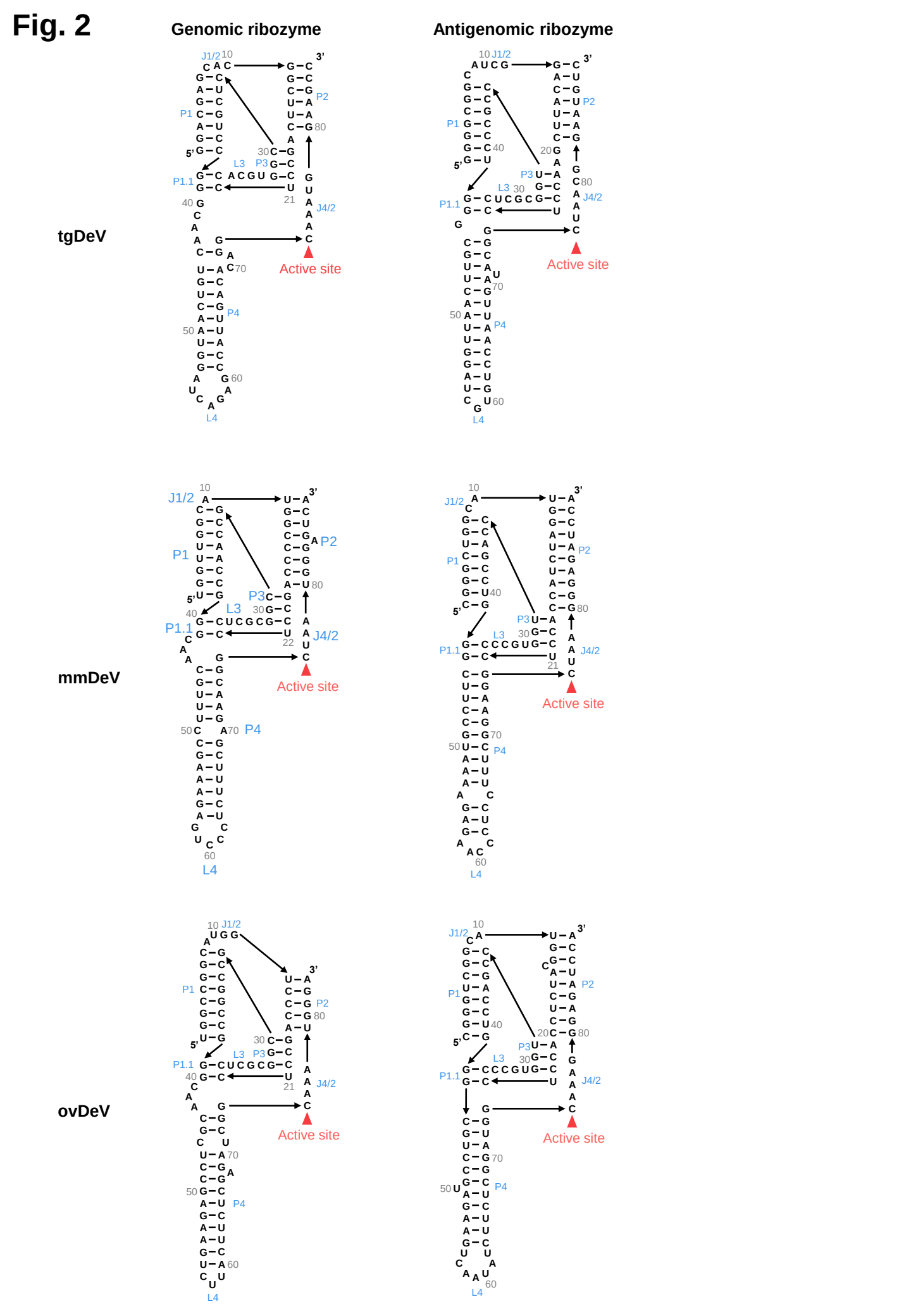
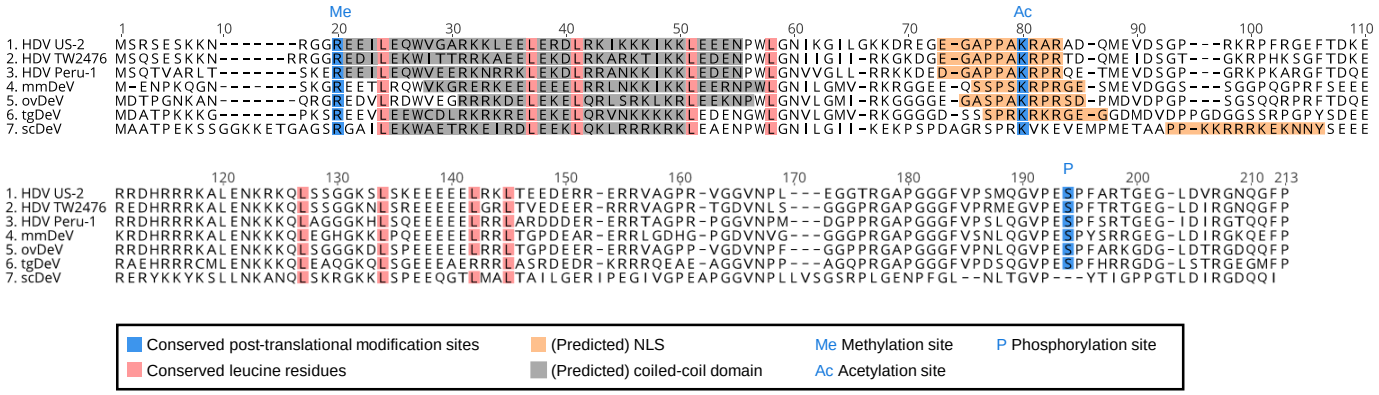
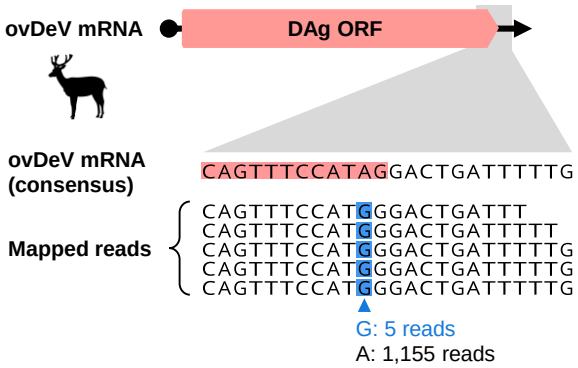

Genomic ribozyme
Antigenomic ribozyme
tgDeV
mmDeV
ovDeV
Active site
J1/2
J4/2
P1
P1.1
P2
P3
P4
L3
L4

Fig. 3

a


Me
Ac
P
1. HDV US-2
2. HDV TW2476
3. HDV Peru-1
4. mmDeV
5. ovDeV
6. tgDeV
7. scDeV
Conserved post-translational modification sites
Conserved leucine residues
(Predicted) NLS
(Predicted) coiled-coil domain
Me Methylation site
Ac Acetylation site
P Phosphorylation site

b


ovDeV mRNA
DAg ORF
ovDeV mRNA (consensus)
CAGTTTCCATAGGACTGATTTTTG
Mapped reads
CAGTTTCCATGGGACTGATTT
CAGTTTCCATGGGACTGATTTTT
CAGTTTCCATGGGACTGATTTTTG
CAGTTTCCATGGGACTGATTTTTG
CAGTTTCCATGGGACTGATTTTTG
G: 5 reads
A: 1,155 reads

c

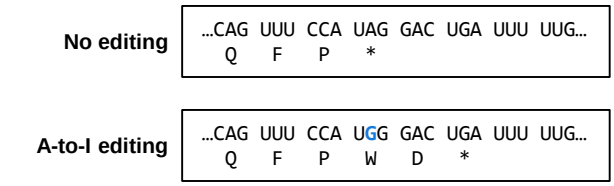

No editing
...CAG UUU CCA UAG GAC UGA UUU UUG...
Q F P *
A-to-I editing
...CAG UUU CCA UGG GAC UGA UUU UUG...
Q F P W D *

Fig. 4

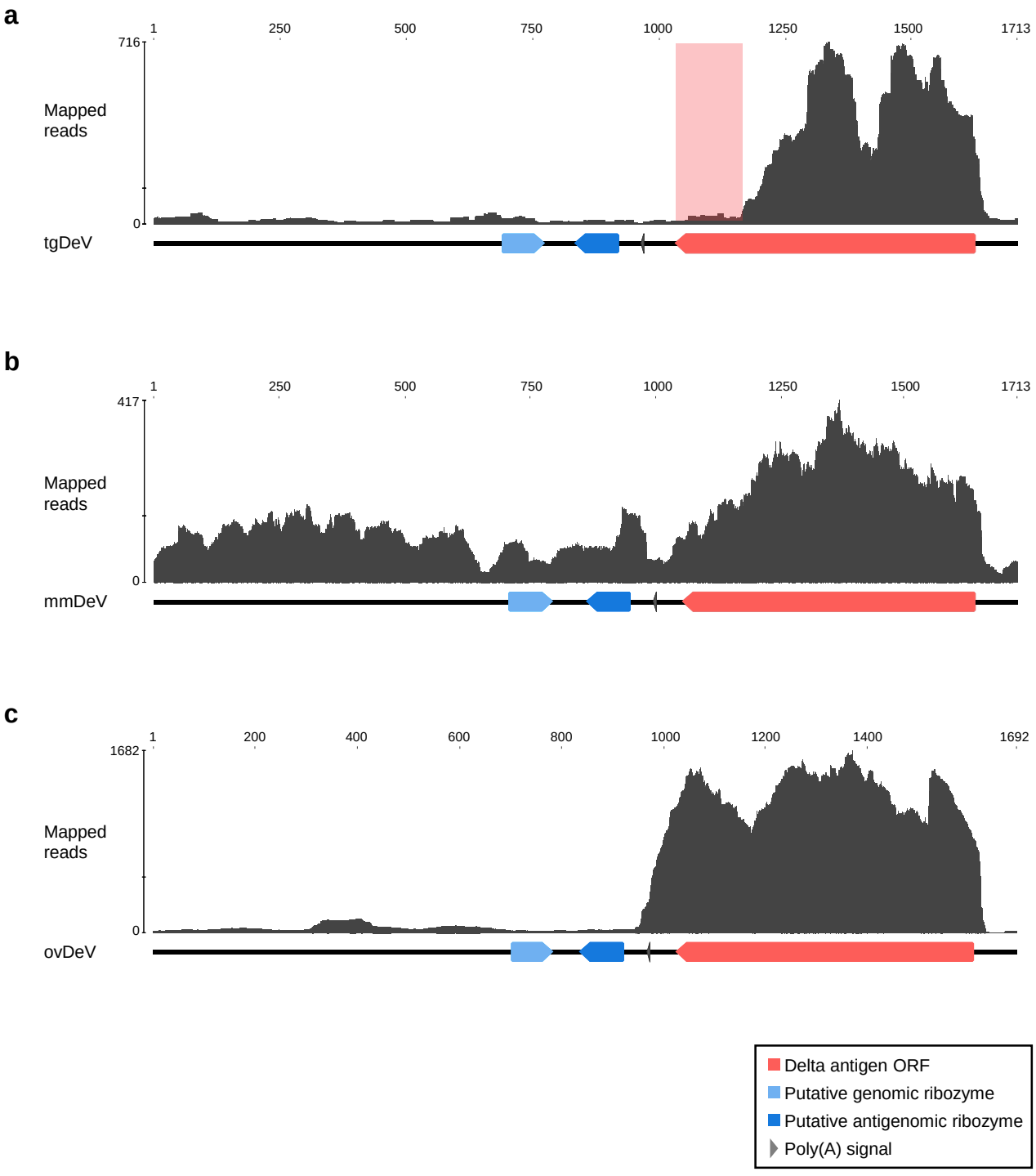

a
1
250
500
750
1000
1250
1500
1713
716
Mapped reads
0
tgDeV
b
1
250
500
750
1000
1250
1500
1713
417
Mapped reads
0
mmDeV
c
1
200
400
600
800
1000
1200
1400
1692
1682
Mapped reads
0
ovDeV
Delta antigen ORF
Putative genomic ribozyme
Putative antigenomic ribozyme
Poly(A) signal

Fig. 5

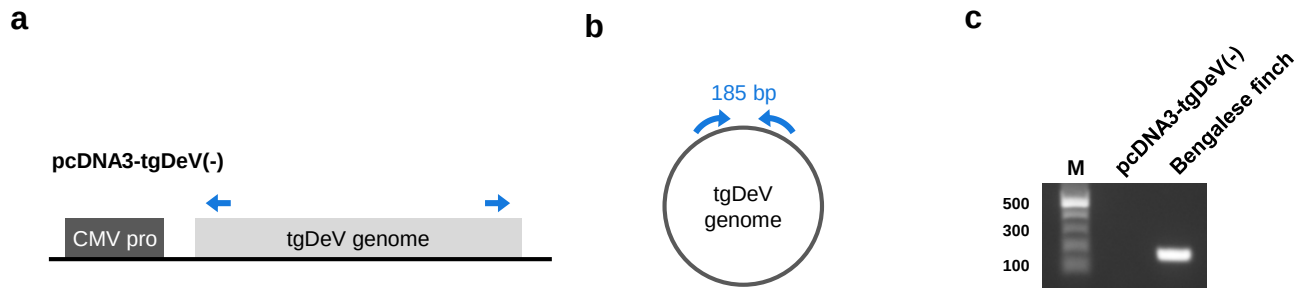
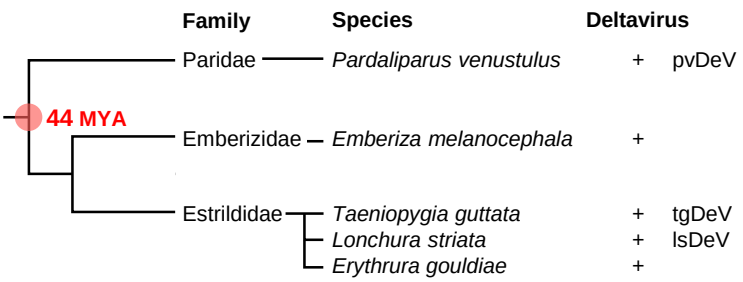

a
pcDNA3-tgDeV(-)
CMV pro
tgDeV genome
b
185 bp
tgDeV
genome
c
M
pcDNA3-tgDeV(-)
Bengalese finch
500
300
100

Fig. 6

a



b

Nucleotide identity (%)

	pvDeV	tgDeV	lsDeV
pvDeV		98.2	98.4
tgDeV	98.2		98.2
lsDeV	98.4	98.2	

Fig. 7

a

	M55042 HDV1	AY261459 HDV2	L22063 HDV3	AF209859 HDV4	AM183328 HDV5	AM183332 HDV6	AM183333 HDV7	AM183330 HDV8	BR001662 ovDeV	BR001661 mmDeV	MK598012 RDeV	BR001665 tgDeV	MH988742 SDeV	MH824555 duck DeV	BR001664 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
BR001662 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
BR001661 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
BR001665 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
BR001664 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